X-RAY CRYSTALLOGRAPHYIC AND BIOCHEMICAL STUDIES OF PSEUDOURIDINE MONOPHOSPHATE GLYCOSIDASE AND THIAMIN PYRIMIDINE SYTNTHASE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Siyu Huang

January 2013

X-RAY CRYSTALLOGRAPHYIC AND BIOCHEMICAL STUDIES OF PSEUDOURIDINE MONOPHOSPHATE GLYCOSIDASE AND THIAMIN PYRIMIDINE SYTNTHASE

Siyu Huang, Ph. D.

Cornell University 2013

Enzymes are macromolecular machines in the cells that catalyze the chemistry of life. Structural and biochemical studies of enzymes and its ligands shed lights on the catalytic mechanism and guide efforts of rational drug design. In this work, three-dimensional structures and biochemical analysis are presented to expand our understanding of two independent enzymes, pseudouridine monophosphate (ΨMP) glycosidase and thiamine pyrimidine synthase.

Pseudouridine (Ψ), the C-isomer of uridine, is the most abundant modified nucleotide in nature. Despite extensive mechanistic study of Ψ biosynthesis, the degradation pathway for Ψ has not been discovered until recently. In this pathway, Ψ is first phosphorylated by Ψ kinase and then degraded into uracil and ribose 5-phosphate by ΨMP glycosidase. Here we present crystallographic and biochemical study of ΨMP glycosidase. Four unique structures of ΨMP glycosidase were displayed to illustrate the reaction pathway of ΨMP degradation. The combination of structural and biochemical study elucidates the mechanism of Ψ degradation and stands as the first example of the mechanistic study of the unusual C-C glycosidic bond cleavage.

Thiamin pyrophosphate is the active form of vitamin B_1 and an essential cofactor for all living systems. The synthesis of thiamin pyrophosphate involves coupling of the thiazole and pyrimidine moieties, which are synthesized separately. All enzymes in prokaryotes and eukaryotes involved in the thiamin biosynthesis pathway have been structurally characterized, except the eukaryotic thiamin pyrimidine synthase THI5p. THI5p synthesizes the pyrimidine moiety of thiamin from histidine and pyridoxal phosphate. This study suggested that THI5p served as the histidine source and became inactive after a single turnover. Moreover, crystal structure of THI5p revealed the identity of the histidine residue and the spatial relationship between the imidazole and the pyridoxal ring. As the reaction also required iron and oxygen, a starting mechanism was proposed, involving Diels-Alder reaction followed by iron-mediated oxidation. Additional structural evidence showed oxidative modification of a cysteine residue at the potential iron binding site, in THI5p after the reaction. The irreversible modification of the cysteine thiol provided further insights into the remarkable chemistry of this suicide enzyme.

BIOGRAPHICAL SKETCH

Siyu was born in Xuzhou, China and is the daughter of Bin Huang and Guanyi Xu. At the age of four, her family moved to Nanjing, an ancient capital of China for six dynasties and settled there since then.

She became curious in science since her elementary school when she participated in the Association of Young Environmental Scientists held by UNESCO. Her passion in science was further intrigued in high school when she studied for the National Biology Olympiad and was fully immersed in the kingdom of plants, animals and the microscopic world of life. With the dream of becoming a scientist in the future, she tapped into the world of chemistry after attending Xiamen University. In the junior year, she went to Lund University in Sweden as an exchange student. She joined Professor Hans-Erik Åkerlund's research group and performed independent research on zeaxanthin, one of the most common carotenoid alcohols found in nature, for the food processing industry. Upon finishing her exchange program, she encountered the Swedish Consortium for Artificial Photosynthesis and was fascinated by their pioneering research in biofuel generation. Without hesitation, she became a member of the consortium and finished her bachelor thesis in Professor Peter Lindblad's group in the Department of Photochemistry and Molecular Science in Uppsala University.

In 2009, Siyu entered Cornell and joined Professor Steven Ealick's group as a doctoral candidate. She carried out research in mechanistic study of enzymes using X-ray crystallography in vitamin and nucleoside metabolic pathways. She has had the

opportunities to work on several interesting enzymes such as pseudouridine monophosphate glycosidase and thiamin pyrimidine synthase.

Apart from her passion in science, Siyu also has diverse interests in art and sports. She plays violin and clarinet since eight and they are still the best friends in her free time. She also loves reading, oil painting and Chinese Calligraphy. She is a member of the Amber Dance Troupe and performed Chinese traditional dance in several ceremonies and festivals in Cornell. She is an avid runner and has completed the marathon in Xiamen. She was an active member of the mountaineering club in her undergraduate university. She was also the Minister of Education in the Student Union where she organized several university-wide cultural festivals and dedicated to the exploration in history, philosophy, art and science.

Thank you my parents,

for giving me such wonderful life,

Thank you my fiancé,

for teaching me how to love and to be loved,

Thank you my friends,

for sharing tears and joys along my journey.

ACKNOWLEDGMENTS

I would like to specially thank Prof. Steve Ealick for mentoring me over the last three and half years and giving me the opportunity to work in the fascinating world of protein crystallography. I am grateful for his guidance that helps me grow as a researcher and a person. I would like to thank Professor Richard Cerione and Brian Crane for serving on my committee and provide insightful advice on my research projects. I would like to thank Professor Tadhg Begley for welcoming me into his lab and collaborating with his students. I would also like to acknowledge the many helpful members of the beam lines at NE-CAT (Argonne National Laboratory) and MacCHESS for their assistance with data collection and processing.

I am lucky to be able to associate myself with the talented and hard working members in the Ealick group. Dr. Yang Zhang and Ms. Leslie Kinsland have provided tremendous support that makes my research considerably more efficient and productive. Dr. Xuan Zhang has been a wonderful friend and resourceful colleague throughout my doctoral research. We have all been there for each other and taught ourselves many tools and issues of protein purification and crystallization. Dr. Michael Fenwick is a dedicated scientist with intense curiosity and brilliant ideas and a great collaborator in the thiamin pyrimidine synthase project. Ms. Megan Sikowitz, Dr. Katherine Hicks and Ms. Megan Kopp are cheerful and companionable friends. They are always very kind and helpful about almost everything in life and science. I am also grateful for Dr. Debamita Paul, Dr. Jarrod French, Dr. Andrew Torelli, Dr. Kathyrn McCulloch, Dr. Timothy Tran and Dr. Shridhar Bale for helping me during

the early stage of my doctoral research and guiding me through the challenges and struggles in protein crystallization and structure determination.

I would also like to thank members in the Begley group for the close collaboration and friendship. I am so grateful for the warm-hearted accommodation during my visit in Texas. Short as it was, I wholeheartedly cherish the precious memory of working with them. Mr. Rung-Yi Lai has been an outstanding collaborator in the thiamin pyrimidine synthase project. He is always there to listen and discuss about everything ranging from protein purification to baseball games. I am also greatly indebted to my mentor in enzyme kinetics, Mr. Nilkamal Mahanta. He is a smart and cheerful person and taught me about the all the kinetic assays and mass spectrometry analysis required in the project of pseudouridine monophosphate glycosidase. I also thank Mr. Angad Mehta, Dr. Lisa Cooper, Ms. Yiquan Liu, Dr. B.J. Philmus, Dr. Dinuka Abeydeera, Dr. Sameh Abdelwahed for your wonderful company and support during my stay in College Station. I will never forget the many delightful lunches, dinners and fun activities we have done together.

I would like to thank Ms. Nattakan Sukomon from Professor Brian Crane's lab. Since we entered Cornell three and half years ago, she has been through a lot with me, celebrating the good times, struggling through the bad times, and being there for all time. I would also like to thank the other members of the Cornell community who provided me with generous support, particularly Dr. Cynthia Kinsland in the Protein Production Facility and Mr. Tom Payne from Prof. Brian Crane's lab and his generous support in the circular dichroism spectra for pseudouridine monophosphate glycosidase.

Last but not least, I thank my family: my parents, Bin Huang, Guanyi Xu, for giving me life in the first place, for educating me with aspects from both arts and science. Your love and support have helped me not only through graduate school, but also my life. My fiancé, Yingchao Yu, for savoring the pain and joy with me over the past seven years, taking care of me through all those harrowing days and sleepless nights, and loved me unconditionally.

TABLE OF CONTENTS

Biographical Sketch iii

Dedication v

Acknowledgements vi

Table of Contents ix

List of Figures xi

List of Tables xiii

List of Abbreviations xiv

Chapter 1. Introduction

Section 1.1. Overview of Structural Biology 1

Section 1.2. Pseudouridine Degradation 2

Section 1.3. Vitamin B_1 Biosynthesis 5

Section 1.4. Acknowledgements 7

References 8

Chapter 2. Pseudouridine Monophosphate Glycosidase: a New Glycosidase Mechanism

Section 2.1. Introduction 13

Section 2.2. Materials and Methods 15

Section 2.3. Results 25

Section 2.4. Discussion 35

References 43

Chapter 3. Thiamin Pyrimidine Biosynthesis in *Candida albicans*: a Remarkable Reaction between Histidine and Pyridoxal Phosphate

Section 3.1. Introduction 49

Section 3.2. Materials and Methods 50

Section 3.3. Results 58

References 68

Chapter 4. Oxidative Modification of Cysteine into Sulfonic Acid in Suicide Enzyme Thiamin Pyrimidine Synthase

Section 4.1. Introduction 71

Section 4.2. Materials and Methods 73

Section 4.3. Results 77

Section 4.4. Discussion 86

References 92

Chapter 5. Conclusion 99

References 103

LIST OF FIGURES

Figure 1.1. Reaction catalyzed by Ψ kinase and ΨMP glycosidase 4

Figure 1.2. Thiamin pyrophosphate biosynthetic pathway in *S. cerevisiae* 5

Figure 2.1. Reactions catalyzed by Ψ kinase and ΨMP glycosidase 10

Figure 2.2. Structure of ΨMP glycosidase 23

Figure 2.3. Stereoview of Mn(II) binding site 24

Figure 2.4. ΨMP glycosidase active site 25

Figure 2.5. Comparison of ΨMP glycosidase complexes 27

Figure 2.6. Analysis of the R5P adduct by mass spectrometry 28

Figure 2.7. Steady state kinetics for ΨMP glycosidase and mutants in the ΨMP formation reaction 31

Figure 2.8. The CD spectra of ΨMP glycosidase and the D149A mutant with and without the addition of $MnCl_2$ 33

Figure 2.9. Mechanism of pseudouridine formation 34

Figure 2.10. Mechanistic proposal for ΨMP glycosidase 35

Figure 3.1. Thiamin pyrophosphate biosynthesis in *S. cerevisiae* 46

Figure 3.2. Reconstitution of the THI5 catalyzed reaction 54

Figure 3.3. MS analysis of HMP-P 55

Figure 3.4. The ESI-MS analysis of ^{14}N-THI5p and ^{15}N-THI5p 56

Figure 3.5. The THI5p denatured in urea and refolded is active 57

Figure 3.6. Sequence alignment for THI5p 58

Figure 3.7. LC-MS analysis of each of the six histidine-containing tryptic peptides derived from active and inactive THI5p 60

Figure 3.8. Active site of the *C. albicans* THI5p showing PLP bound via an imine to Lys62 and His66 in close proximity to the PLP 61

Figure 3.9. Mechanistic proposal for the formation of HMP-P from PLP and the active

site His66 62

Figure 4.1. Overall structure of THI5p 74

Figure 4.2. Active site of THI5p/PLP complex. 76

Figure 4.3. Difference Distance Matrix Analysis of THI5p/PLP complex vs. modified-THI5p 77

Figure 4.4. Composite omit map for residues in modified-THI5p 79

Figure 4.5. The structural comparisons of THI5p/PLP complex with the modified-THI5p, the C199A/PLP complex, the E309Ter/PLP complex and the THIYp/FAMP complex 80

Figure 4.6. Analysis of Cys199 by LC-MS 82

Figure 4.7. Active site of THI5p with PLP removed, and FAMP modeled in based on the superposition of the structures of THI5p/PLP and THIYp/FAMP. 85

LIST OF TABLES

Table 2.1. Primer sequences of ΨMP glycosidase and the mutants 15

Table 2.2. Data collection statistics 18

Table 2.2. Data refinement statistics 19

Table 2.3. Steady state kinetic parameters for ΨMP glycosidase and mutants 32

Table 3.1. Data collection statistics 52

Table 3.2. Data refinement statistics 53

Table 3.3. THI5p histidine mutants and their activities 59

Table 4.1. Data collection statistics 71

Table 4.2. Data refinement statistics 72

LIST OF ABBREVIATIONS

APS, Advanced Photon Source

CD, Circular dichroism

CHESS, Cornell High Energy Synchrotron Source

Cryo-EM, Cryo-electron microscopy

NMR, Nuclear Magnetic Resonance Spectroscopy

DTT, Dithiothreitol

EcΨMP glycosidase, *Escherichia coli* ΨMP glycosidase

FAMP, N-formyl-4-amino-5-(aminomethyl)-2-methylpyrimidine

FADH2, Flavin adenine dinucleotide, reduced

HMP, 2-Methyl-4-amino-5-hydroxymethylpyrimidine

HMP-P, 2-Methyl-4-amino-5-hydroxymethylpyrimidine phosphate

LC, Liquid Chromatography

LB, Luria-Bertani

MS, Mass spectrometry

NE-CAT, Northeast Collaborative Access Team

NAD, Nicotinamide adenine dinucleotide

PLP, pyridoxal 5′-phosphate

PBP, periplasmic binding protein

ΨMP, Pseudouridine 5′-monophosphate

Ψ, Pseudouridine

RMSD, Root mean square deviation

R5P, Ribose 5-phosphate

TCEP, Tris(2-carboxyethyl)phosphine

TmΨMP glycosidase, *Thermotoga maritima* ΨMP glycosidase

Tris, tris(hydroxymethyl)aminomethane

CHAPTER 1

INTRODUCTION

Section 1.1. Overview of Structural Biology

Proteins are a diverse and abundant family of biopolymers. Comprising about 50% of a cell's dry weight, proteins are major players in almost every essential process of life, including serving as basic building blocks, catalyzing a multitude of biochemical reactions, and maintaining a stable cellular environment. The basic monomer of a protein is the amino acid. The primary structure of protein is simply the arrangement of amino acids in a single polypeptide chain. Interactions between adjacent amino acids reshape the polypeptide chain into secondary structures such as helical or pleated segments. The secondary structures of proteins further fold into globular tertiary structures. Some proteins contain multiple polypeptides with characteristic tertiary structures. The organization of tertiary structures constitutes the quaternary structure of proteins.

Recent developments in structural biology based on the characterization of protein structures have tremendously contributed to the understanding of the complex biological system that governs cellular function. By virtue of the availability of multi-wavelength synchrotron sources (*1*), high-resolution crystal structure determination became possible and vast majority of protein structures are determined using X-ray crystallography in the past decade. The advances of computer science and physics also considerably accelerated the process of structure determination. Coupled with kinetic characterization, site-directed mutagenesis, mass spectrometry analysis and

various spectroscopic methods, crystallography has become a powerful tool to study the protein structure-function relationship.

One of the major goals of structural biology is to understand the catalytic mechanisms of enzymes that are important to the cellular processes. As many enzymes are potential drug targets, the elucidation of the mechanistic details of the enzyme can provide valuable information for the rational design of drugs (*2*). X-ray crystallography has been a powerful tool in both the *de novo* design and lead optimization in drug development. For example, the development of the drug Agenerase® (*3*) against HIV protease and Thymitaq® (*4*) for the liver cancer treatment are classic examples of structure-based drug design. The 2012 Nobel Prize in Chemistry was awarded to Dr. Robert Lefkowitz and Dr. Brian Kobilka for their structural and functional studies on G-protein coupled receptors, which are responsible for the majority of transmembrane signal transduction and the are target of about 30% of all drugs (*5*).

Once a drug target has been revealed and elucidated with a structure, leads can be either designed *de novo* or identified through database search. If the synthesis is feasible, a small library of compounds similar to the proposed ligand will be chemically synthesized. Then the compound can be tested in a biochemical or cell-based assay to determine if the design is successful. The initial structure determination of the drug target, the design of the lead, the chemical synthesis of the lead and the series of biological assays constitute a typical cycle of drug design. It is often necessary to go through several cycles of lead optimization before a drug candidate emerges (*3*).

Section 1.2. Pseudouridine Degradation

Among the many dynamic research in the RNA world, a large number of structurally diverse modifications on the polyribonucleotides have be identified and become an active field of research in the past several years. Pseudouridine (Ψ), the C-glycoside isomer of uridine, is not only the first modified nucleotide to be identified, but also the most abundant RNA modification in nature, thus is often termed as the "fifth nucleoside" (*6*). Ψ ubiquitously exists in rRNA, tRNA, snRNA and snoRNA and is formed through site-specific post-transcriptional modification of uridine (*7*). The binding of uracil to ribose through C5 rather than N1 exposes an additional N-bound hydrogen that enables a structured water molecule to coordinate both Ψ and the sugar-phosphate. This rigidifying effect stabilizes the phosphodiester bonds in vicinity (*8-10*), and thereby enhances the stacking of neighboring nucleosides. The stabilizing effect underlies most of the biological functions of Ψ residue in RNA. For example, Ψ is not only found as the nearly universal Ψ55 in the TΨC stem in tRNAs, but also frequently exists in the D stem and the anticodon stem and loop (*11-13*). In tRNAs, during translation, Ψ may be involved in modulating the interactions that tRNA molecules makes with rRNAs and with mRNAs by affecting the local structure of neighboring domains (*14*). Pseudouridylation is also found to play an important role in influencing the decoding activity of tRNAs and improving the fidelity of protein biosynthesis (*15*). In rRNAs, Ψ is believed to involve both rRNA folding and ribosome assembly (*16*), which can be attributed to the structural stability afforded by Ψ. In the mature ribosome, Ψ may contribute to stabilizing local structures through RNA-RNA and RNA-protein interactions (*17*).

Since the Ψ level increases in tumor bearing patients (*18, 19*), the biosynthesis of Ψ has drawn considerable attention in the past few years. A combination of biochemical, biophysical and structural study has been carried out to elucidate the mechanism for Ψ biosynthesis. Studies showed that the biosynthesis of Ψ is carried out by Ψ synthase through a site-specific modification from uridine to Ψ in RNA (*20*). Ψ synthase acts on the maturing RNAs by severing the C-N glycosidic bond, flipping the base along the N3-C6 axis and reattaching the uracil to the ribose through a C-C glycosidic bond (*9*). The C-C glycosidic bond is unusual in nature. Little was known about the degradation pathway of Ψ and the cleavage of C-C glycosidic bond until the recent discovery of two genes that encode two enzymes that catalyze the degradation of Ψ (*21*). Ψ kinase and ΨMP glycosidase are associated in the same operon in bacteria and exist as a homologous bifunctional protein in lower eukaryotes. Human beings do not have the gene that encodes the bifunctional enzyme, and therefore are unable to utilize Ψ. In the Ψ degradation pathway, Ψ is first converted to pseudouridine 5′-monophosphate (ΨMP) by Ψ kinase and then ΨMP is degraded by ΨMP glycosidase to uracil and ribose 5-phosphate (Figure 1.1). ΨMP glycosidase is the first example of a mechanistically and structurally characterized enzyme that cleaves a C-C glycosidic bond (*22*). While O-C and N-C glycosidic bond are prevalent in nature and the mechanisms for cleavage have been extensively studied (*23-26*), the mechanism for C-C glycosidic bond remains unknown. Compared to O-N and C-N glycosidic bond, C-C glycosidic bond is much more stable thermodynamically and resistant towards hydrolysis. Chapter 2 outlines the structural

and biochemical characterizations of *Escherichia coli* ΨMP glycosidase. A series of crystal structures of enzyme-substrate and enzyme-intermediate complexes have been determined. Four unique snapshots are displayed along the reaction coordinate, thus revealing the unprecedented chemistry and expanding our knowledge in glycosidic bond cleavage.

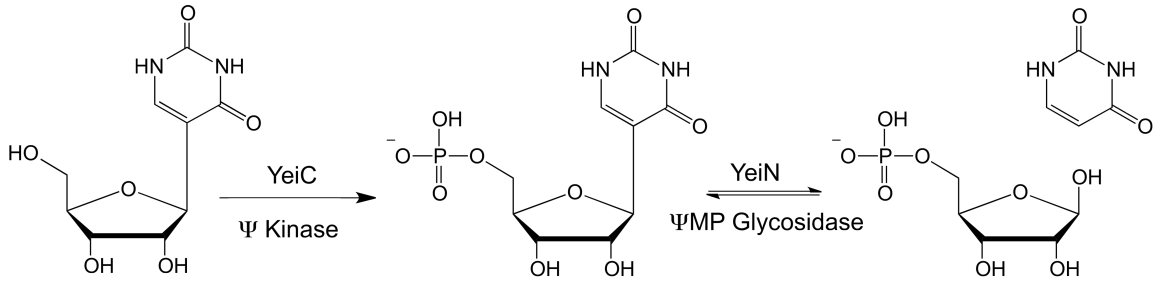


Figure 1.1. Reaction catalyzed by Ψ kinase and ΨMP glycosidase.

Section 1.3. Vitamin B_1 Biosynthesis

Vitamin B_1 is an essential cofactor in all living systems and involved in amino acid synthesis and carbohydrate metabolism (*27*). Vitamin B_1 is found in high concentration in brain, heart, skeletal muscle, kidney and liver in human beings. The deficiency of vitamin B_1 can cause fatal disease such as neurological disorder and cardiac disorder, especially for alcoholics (*28*). Bacteria and archea, as well as low eukaryotes, synthesize thiamine, whereas higher eukaryotes like human beings cannot produce thiamin and require it as dietary supplement. Deficiency in vitamin B_1 in humans causes both cardiovascular system disorder (*29*) and brain disorder (*28*). The transport form of vitamin B_1 is thiamin, while the active form is thiamin pyrophosphate. Thiamin pyrophosphate consists of thiazole and pyrimidine moieties, which are biosynthesized separately in both prokaryotes and eukaryotes (Figure 1.2)

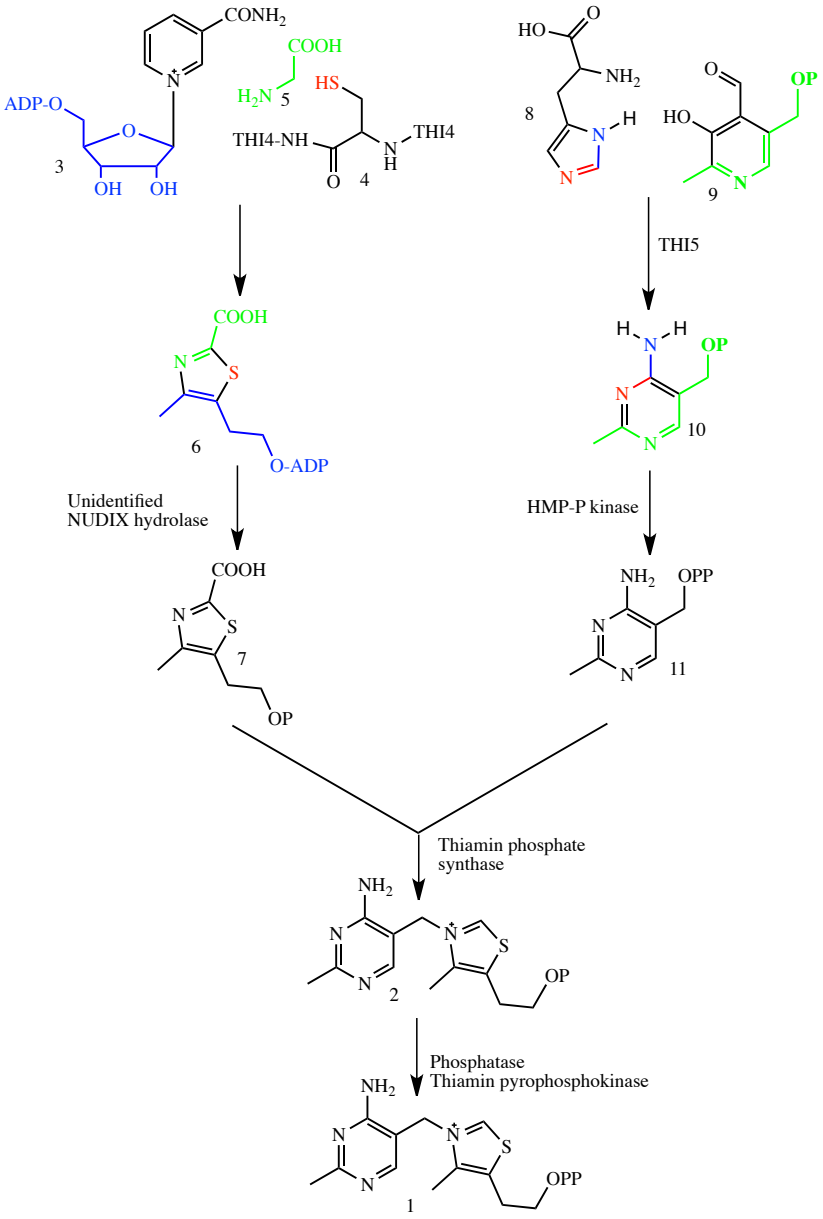


Figure 1.2. Thiamin pyrophosphate biosynthetic pathway in *Saccharomyces cerevisiae*.

and then coupled (*30*). In eukaryotes, thiamine pyrimidine is biosynthesized from histidine and pyridoxal 5′-phosphate (PLP) and utilizes the enzyme THI5p (*31-34*). Chapter 3 demonstrates the reconstitution of activity and preliminary characterization

of the *Candida albicans* pyrimidine synthase (THI5p). A structure of the *C. albicans* THI5p shows a PLP bound in the active site via an imine with Lys62 and His66 is in close proximity to the PLP. The biochemical assay and the crystal structure suggest that His66 is the histidine source for pyrimidine formation and that the pyrimidine synthase is a suicide enzyme that becomes inactive after a single-turnover. Chapter 4 provides further insights into THI5p as a suicide enzyme by determination of the crystal structure of a modified THI5p that undergoes irreversible trioxidation on a cysteine residue at the active site. A series of conformational changes have been observed with local movements in the small domain of the enzyme. The crystal structures of C199A/PLP complex and the C-terminus truncated mutant E309Ter/PLP complex are also determined to reveal the structural evidence of the flexible C-terminus. Along with the support from the LC-MS data, the four structures demonstrate the unique chemistry of this single-turnover enzyme and unveil the remarkable reaction that is without chemical and biochemical precedent.

Section 1.4. Acknowledgements

The work presented in following chapters relies heavily on collaborative efforts. The kinetic study and mass spectrometry analysis on ΨMP glycosidase was performed together with Mr. Nilkamal Mahanta from Professor Tadhg Begley's research group. The reconstitution of activity, mass spectroscopic analysis and biochemical assays of thiamin pyrimidine synthase was carried out by Mr. Rung-Yi Lai from Professor Tadhg Begley's research group. The circular dichroim spectra were taken under the assistance of Mr. Tom Payne from Professor Brian Crane's laboratory.

REFERENCES

1. Phillips, J. C., Wlodawer, A., Yevitz, M. M., and Hodgson, K. O. (1976) APPLICATIONS OF SYNCHROTRON RADIATION TO PROTEIN CRYSTALLOGRAPHY - PRELIMINARY RESULTS, *Proc. Natl. Acad. Sci. U. S. A. 73*, 128-132.
2. Greer, J., Erickson, J. W., Baldwin, J. J., and Varney, M. D. (1994) Application of the three-dimensional structures of protein target molecules in structure-based drug design, *J Med Chem 37*, 1035-1054.
3. Blundell, T. L., Jhoti, H., and Abell, C. (2002) High-throughput crystallography for lead discovery in drug design, *Nat. Rev. Drug Discov. 1*, 45-54.
4. Tong, Y. Z., Liu-Chen, X. Y., Ercikan-Abali, E. A., Capiaux, G. M., Zhao, S. C., Banerjee, D., and Bertino, J. R. (1998) Isolation and characterization of Thymitaq (AG337) and 5-fluoro-2-deoxyuridylate-resistant mutants of human thymidylate synthase from ethyl methanesulfonate-exposed human sarcoma HT1080 cells, *J. Biol. Chem. 273*, 11611-11618.
5. Baker, E. N., and Dauter, Z. (2012) Nobel Prize for Chemistry 2012: GPCRs seen through a crystal ball, *Acta Crystallogr D Biol Crystallogr 68*, 1439-1440.
6. Cohn, W. E. (1960) Pseudouridine, a Carbon-Carbon Linked Ribonucleoside in Ribonucleic Acids - Isolation, Structure, and Chemical Characteristics, *J. Biol. Chem. 235*, 1488-1498.
7. Charette, M., and Gray, M. W. (2000) Pseudouridine in RNA: What, Where, How, and Why, *IUBMB Life 49*, 341-351.
8. Gu, X. R., Liu, Y. Q., and Santi, D. V. (1999) The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA, *Proc. Natl. Acad. Sci. U.S.A. 96*, 14270-14275.

9. Hoang, C., and Ferré-D'Amaré, A. R. (2001) Cocrystal Structure of a tRNA [Psi]55 Pseudouridine Synthase: Nucleotide Flipping by an RNA-Modifying Enzyme, *Cell 107*, 929-939.
10. McDonald, M. K., Miracco, E. J., Chen, J., Xie, Y., and Mueller, E. G. (2010) The Handling of the Mechanistic Probe 5-Fluorouridine by the Pseudouridine Synthase TruA and Its Consistency with the Handling of the Same Probe by the Pseudouridine Synthases TruB and RluA, *Biochemistry 50*, 426-436.
11. Wrzesinski, J., Nurse, K., Bakin, A., Lane, B. G., and Ofengand, J. (1995) A dual-specificity pseudouridine synthase: an Escherichia coli synthase purified and cloned on the basis of its specificity for psi 746 in 23S RNA is also specific for psi 32 in tRNA(phe), *RNA 1*, 437-448.
12. Massenet, S., Motorin, Y., Lafontaine, D. L., Hurt, E. C., Grosjean, H., and Branlant, C. (1999) Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA, *Mol Cell Biol 19*, 2142-2154.
13. Becker, H. F., Motorin, Y., Planta, R. J., and Grosjean, H. (1997) The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs, *Nucleic Acids Res 25*, 4493-4499.
14. Davis, D. R., and Poulter, C. D. (1991) 1H-15N NMR studies of Escherichia coli tRNA(Phe) from hisT mutants: a structural role for pseudouridine, *Biochemistry 30*, 4223-4231.
15. Foster, P. G., Huang, L. X., Santi, D. V., and Stroud, R. M. (2000) The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I, *Nat. Struct. Biol. 7*, 23-27.

16. Cunningham, P. R., Richard, R. B., Weitzmann, C. J., Nurse, K., and Ofengand, J. (1991) The absence of modified nucleotides affects both in vitro assembly and in vitro function of the 30S ribosomal subunit of Escherichia coli, *Biochimie 73*, 789-796.

17. Ofengand, J., Bakin, A., Wrzesinski, J., Nurse, K., and Lane, B. G. (1995) The pseudouridine residues of ribosomal RNA, *Biochem Cell Biol 73*, 915-924.

18. Feng, B., Zheng, M. H., Zheng, Y. F., Lu, A. G., Li, J. W., Wang, M. L., Ma, J. J., Xu, G. W., Liu, B. Y., and Zhu, Z. G. (2005) Normal and modified urinary nucleosides represent novel biomarkers for colorectal cancer diagnosis and surgery monitoring, *J. Gastroen. Hepatol. 20*, 1913-1919.

19. Seidel, A., Brunner, S., Seidel, P., Fritz, G. I., and Herbarth, O. (2006) Modified nucleosides: an accurate tumour marker for clinical diagnosis of cancer, early detection and therapy control, *Brit. J. Cancer 94*, 1726-1733.

20. Hamilton, C. S., Greco, T. M., Vizthum, C. A., Ginter, J. M., Johnston, M. V., and Mueller, E. G. (2006) Mechanistic Investigations of the Pseudouridine Synthase RluA Using RNA Containing 5-Fluorouridine†, *Biochemistry 45*, 12029-12038.

21. Preumont, A., Snoussi, K., Stroobant, V., Collet, J. F., and Van Schaftingen, E. (2008) Molecular identification of pseudouridine-metabolizing enzymes, *J. Biol. Chem. 283*, 25238-25246.

22. Huang, S., Mahanta, N., Begley, T. P., and Ealick, S. E. (2012) Pseudouridine Monophosphate Glycosidase: A New Glycosidase Mechanism, *Biochemistry 51*, 9245-9255.

23. Gopaul, D. N., Meyer, S. L., Degano, M., Sacchettini, J. C., and Schramm, V. L. (1996) Inosine-uridine nucleoside hydrolase from *Crithidia fasciculata*.

Genetic characterization, crystallization, and identification of histidine 241 as a catalytic site residue, *Biochemistry 35*, 5963-5970.

24. Duerre, J. A. (1962) Hydrolytic Nucleosidase Acting on *S*-Adenosylhomocysteine and on 5'-Methylthioadenosine, *J. Biol. Chem. 237*, 3737-&.

25. Hurwitz, J., Heppel, L. A., and Horecker, B. L. (1957) The Enzymatic Cleavage of Adenylic Acid to Adenine and Ribose 5-Phosphate, *J. Biol. Chem. 226*, 525-540.

26. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms, *Annu Rev Biochem 77*, 521-555.

27. Lonsdale, D. (2006) Review of the biochemistry, metabolism and clinical benefits of thiamin(e) and its derivatives, *Evid.-based Complement Altern. Med. 3*, 49-59.

28. Harper, C. (2006) Thiamine (vitamin B1) deficiency and associated brain damage is still common throughout the world and prevention is simple and safe!, *Eur J Neurol 13*, 1078-1082.

29. Weiss, S. (1940) Occidental beriberi with cardiovascular manifestations - Its relation to thiamin deficiency, *J. Am. Med. Assoc. 115*, 832-839.

30. Jurgenson, C. T., Begley, T. P., and Ealick, S. E. (2009) The structural and biochemical foundations of thiamin biosynthesis, *Annu Rev Biochem 78*, 569-603.

31. Zeidler, J., Ullah, N., Gupta, R. N., Pauloski, R. M., Sayer, B. G., and Spenser, I. D. (2002) 2'-hydroxypyridoxol, a biosynthetic precursor of vitamins B(6) and B(1) in yeast, *J Am Chem Soc 124*, 4542-4543.

32. Zeidler, J., Sayer, B. G., and Spenser, I. D. (2003) Biosynthesis of vitamin B1 in yeast. Derivation of the pyrimidine unit from pyridoxine and histidine. Intermediacy of urocanic acid, *J Am Chem Soc 125*, 13094-13105.

33. Tazuya, K., Azumi, C., Yamada, K., and Kumaoka, H. (1995) Pyrimidine moiety of thiamin is biosynthesized from pyridoxine and histidine in Saccharomyces cerevisiae, *Biochem Mol Biol Int 36*, 883-888.

34. Ishida, S., Tazuya-Murayama, K., Kijima, Y., and Yamada, K. (2008) The direct precursor of the pyrimidine moiety of thiamin is not urocanic acid but histidine in Saccharomyces cerevisiae, *J Nutr Sci Vitaminol (Tokyo) 54*, 7-10.

CHAPTER 2

PSEUDORURIDINE MONOPHOSPHATE GLYCOSIDASE: A NEW GLYCOSIDASE MECHANISM[1]

Section 2.1. Introduction

Pseudouridine (Ψ) was the first modified nucleoside to be discovered and is the most abundant modification in RNA, existing in tRNAs, rRNAs, snRNAs and snoRNAs (*1*). Ψ is biosynthesized post-transcriptionally from RNA uridine moieties by Ψ synthase, which cleaves the glycosidic C-N bond and reconnects the uracil to the ribosyl moiety at the C5 position (*2*). Pseudouridine reduces the conformational flexibility of RNA because the exposed, protonated N1 atom forms strong hydrogen bond with a structured water molecule, replacing weak interactions formed by C5 (*1, 3*).

While Ψ biosynthesis has been studied extensively, little is known about Ψ catabolism. Recently, the two enzymes responsible for Ψ **1** degradation in *Escherichia coli* were identified (*4*). Ψ kinase first phosphorylates Ψ to pseudouridine 5′-phosphate (ΨMP, **2**). ΨMP glycosidase then catalyzes the reversible cleavage of the C-C glycosidic bond to form uracil **4** and ribose 5-phosphate (R5P, **3**) (Figure 2.1).

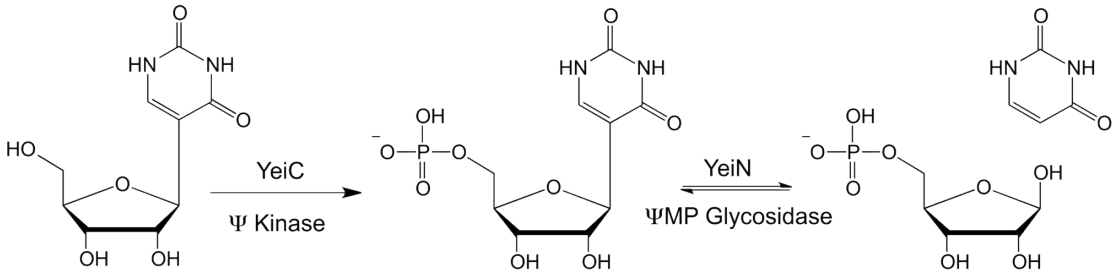


Figure 2.1. Reactions catalyzed by Ψ kinase and ΨMP glycosidase.

---

[1] Produced with permission from Huang S., Mahanta N., Begley T. P., Ealick S. E., (2012), *Biochemistry*. Copyright © 2012 American Chemical Society

While ΨMP glycosidase functions biologically in the cleavage direction, equilibrium strongly favors ΨMP synthesis with an equilibrium constant for the degradation reaction of 2.3×10^{-4} M (*4*). In bacterial genomes Ψ kinase and ΨMP glycosidase are encoded by separate genes. In *E. coli* these genes are *yeiC* and *yeiN*, respectively, which are located in the same operon. In some eukaryotes, the genes for Ψ kinase and ΨMP glycosidase are fused, resulting in a bifunctional enzyme. Enzymes metabolizing Ψ are not found in humans and most other higher organisms, and excess Ψ is excreted (*5*).

While natural products containing O and N glycosidic bonds are widespread, natural products containing the C-glycosidic bond are much less prevalent (*6*). The C-glycosidic bond is generally assumed to be formed by an electrophilic aromatic substitution. Enzymes that cleave C-N glycosidic bonds have been extensively studied. These include nucleoside and nucleotide hydrolases (*7-9*), purine and pyrimidine nucleoside phosphorylases (*10*), purine and pyrimidine phosphoribosyltransferases (*11, 12*), and nucleoside deoxyribosyltransferases (*13*). In addition, DNA-repair enzymes such as endonuclease III function as N-glycosylases to excise damaged nucleobases from DNA (*14*). Enzymes that cleave C-O glycosidic bonds are widespread in carbohydrate metabolism and have also been extensively studied (*15*). Biochemical, biophysical and genetic studies suggest that these enzymes cleave C-N and C-O glycosidic bonds by a largely dissociative mechanism generating an oxocarbenium ion intermediate. While the genes encoding several C-glycosynthases have been identified and in many cases overexpressed (*16-23*) and the structures of ligand free UrdGT2 (*24, 25*) and ΨMP glycosidase (*26*) have been determined – there is currently no example of a mechanistically well-characterized C-glycosidase or glycosynthase.

Here we report a set of *E. coli* ΨMP glycosidase (EcΨMP glycosidase) crystal structures, which provide four snapshots of the reaction coordinate. The structures are: the unliganded enzyme, a ΨMP glycosidase/ring-opened R5P adduct, a ΨMP glycosidase/ ring-opened ribose ΨMP adduct, and a K166A ΨMP complex. The structural studies suggested that the reaction involves an intermediate imine with Lys166. This was confirmed by mass spectrometry. Kinetic studies on active site mutants suggested roles for individual amino acid side chains. Comparison of the EcΨMP glycosidase structures suggested a role for conformational changes in ΨMP cleavage and product release. These results suggest an unanticipated mechanism for the ΨMP glycosidase catalyzed C-glycosyl bond hydrolysis.

Section 2.2. Materials and Methods

Cloning, Overexpression and Purification of ΨMP Glycosidase. Standard methods were used for DNA restriction endonuclease digestion, ligation and transformation of DNA (*27*). Automated DNA sequencing was performed at the Cornell BioResource Center. Plasmid DNA was purified with a GeneJet miniprep kit (Fermentas, Glen Burnei, MD). DNA fragments were separated by agarose gel electrophoresis, excised and purified with the Zymoclean gel DNA recovery kit (Zymo Research, Orange, CA). *E. coli* strain MachI (Invitrogen, Madison, WI) was used as the recipient for transformations during plasmid construction and for plasmid propagation and storage. An Eppendorf Mastercycler and Phusion DNA polymerase (New England Biolabs, Ipswich, MA) were used for PCR. All restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (Ipswich, MA). *E. coli* strain BL21(DE3) and the pET overexpression system were purchased from Novagen (Madison, WI).

The *yeiN* gene was PCR amplified from *E. coli* K12 genomic DNA using the following primers: upstream primer 5′-GGG TAG CAT ATG TCT GAA TTA AAA

ATT TCC CCT G-3′ (inserts an *Nde*I site at the start codon of the *yeiN* open reading frame); downstream primer 5′-CCC TAC TCG AGT TAA CCC GCG AGA CGC TGA TAT TC-3′ (inserts an *Xho*I site after the end of the *yeiN* open reading frame). The purified PCR product was digested with *Nde*I and *Xho*I, purified and ligated into similarly digested pTHT, a pET-28 derived vector, which allows attachment of a modified six histidine tag followed by a tobacco etch virus protease cleavage site onto the N-terminus of the expressed protein. Colonies were screened for the presence of the insert and a representative plasmid was designated pEcYeiN.THT. The PCR-derived DNA was sequenced and shown to contain no errors.

The *yeiN* gene was further transformed into *E. coli* strain BL21(DE3) (Novagen). The cells were grown overnight in a 10 mL starter culture in Luria-Bertani (LB) media (*28*) containing 30 μg/mL kanamycin, then transferred to cultures containing 1.5 L LB media and incubated at 37 ºC with shaking until an OD_{600} of 0.8 was achieved. The culture was then induced with 1 mM isopropyl-1-β-D-galactopyranoside and incubated overnight at 15 ºC. Cells were harvested by centrifugation at 10000 *g* for 30 min at 4 ºC and resuspended in a lysis buffer containing 50 mM tris(hydroxymethyl)aminomethane (Tris), pH 8.0, 300 mM NaCl, 10 mM imidazole, and lysed by sonication. The lysate was centrifuged at 40000 *g* for 30 min and the supernatant was loaded onto a column containing 2 mL Ni-NTA resin (Qiagen) preequiliburated with the lysis buffer. The column was then washed with three times the volume of wash buffer containing 50 mM Tris, pH 8.0, 300 mM NaCl and 30 mM imidazole for 1.5 h. The protein was eluted with 50 mM Tris, pH 8.0, 300 mM NaCl, and 250 mM imidazole. The eluted protein was then subjected to size exclusion chromatography using an ACTA Explorer FPLC with a HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare). The resulting protein was more than 95% pure as judged by SDS-PAGE analysis (unpublished experiments). The

protein was then concentrated to 15 mg/mL using an Amicon concentrator (30 kDa MWCO filter, Millipore), flash frozen and stored at -80 ºC.

ΨMP Glycosidase Mutagenesis. Site directed mutagenesis was performed by a standard PCR protocol using P*fu*UltraII DNA polymerase per the manufacturer's instructions (Agilent) and *Dpn*I to digest the methylated parental DNA prior to transformation. In addition to the forward and reverse primers required to introduce the mutation, a third primer was designed to screen for the presence of the mutation by colony PCR (Table 2.1). For screening, the primers designated "sF" were paired with the T7T primer (5′-GCTAGTTATTGCTCAGCGG-3′) and the primers designated "sR" were screened with the T7Plac primer (5′-TATAGGGGAATTGTGAGCGG-3′). All the mutants were verified by sequencing.

Enzymatic Synthesis of ΨMP. 2.5 mg ΨMP glycosidase, 100 mM uracil and R5P were incubated in a buffer containing 0.5 mM $MnCl_2$, 25 mM HEPES, pH 7.1 at 25 °C for 1 h. The enzyme was removed using an Amicon 30 kDa MWCO (Millipore) filter. The reaction product was identified as ΨMP by comigration with an authentic sample of ΨMP during HPLC analysis .

Crystallization of ΨMP Glycosidase. ΨMP glycosidase was crystallized using the hanging drop vapor diffusion method at 22 ºC. The initial crystallization condition was determined using sparse matrix screens Crystal Screen 1 and 2 (Hampton Research). The optimized reservoir conditions were 20% polyethylene glycol 4000, 0.2 M sodium acetate and 0.1 M Tris, pH 7.0. ΨMP glycosidase was incubated with 4 mM $MnSO_4$ before crystallization. The drops contained 1.5 µL of protein solution and 1.5 µL of reservoir solution. Prismatic crystals grew within 2 d to the size of 200 µm × 50 µm × 50 µm.

The ΨMP glycosidase/R5P complex was co-crystallized under similar conditions except that the protein was incubated with 4 mM $MnCl_2$ and 2 mM R5P for

Table 2.1. Primer sequences of ΨMP glycosidase and the mutants. In addition to the forward and reverse primers required to introduce the mutation, a third primer was designed to screen for the presence of the mutation by colony PCR. The primers designnated "sF" were paired with the T7T primer (5'-GCTAGTTATTGCTCAGCGG-3') and the primers designated "sR" were screened with the T7Plac primer (5'-TATAGGGGAATTGTGAGCGG-3').

Mutant	Primer Sequences*	
D149A	D149A-F	ccttcgatatttctgccgCGttgcaagaactggcaaatac
	D149A-R	gtatttgccagttcttgcaaCGcggcagaaatatcgaagg
	D149A-sF	ccttcgatatttctgccgCG
E31A	E31A-F	caaaaaaccggttgtggcgctggCGtcgaccattatttctcac
	E31A-R	gtgagaaataatggtcgaCGccagcgccacaaccggttttttg
	E31A-sF	aaccggttgtggcgctggCG
E31L	E31L-F	caaaaaaccggttgtggcgctgCTGtcgaccattatttctcac
	E31L-R	gtgagaaataatggtcgaCAGcagcgccacaaccggttttttg
	E31L-sF	aaccggttgtggcgctgCTG
K93A	K93A-F	cgtgaagggcataacgtgaccGCGgttagtcgtcgcgatttacc
	K93A-R	ggtaaatcgcgacgactaacCGCggtcacgttatgcccttcacg
	K93A-sF	aagggcataacgtgaccGCG
K93E	K93E-F	cgtgaagggcataacgtgaccGaGgttagtcgtcgcgatttacc
	K93E-R	ggtaaatcgcgacgactaacCtCggtcacgttatgcccttcacg
	K93E-sF	aagggcataacgtgaccGaG
K166A	K166A-F	ccgttgtttgtgccggggcgGCGtctattctcgatttagg
	K166A-R	cctaaatcgagaatagaCGCcgccccggcacaaacaacgg
	K166A-sF	ttgtttgtgccggggcgGCG
H137A	H137A-F	ccgggggaattggtggtgtgGCGcgcggggcggaacataccttcg
	H137A-R	cgaaggtatgttccgccccgcgCGCcacaccaccaattcccccgg
	H147A-sF	ggggaattggtggtgtgGCG

30 min prior to crystallization. The ΨMP glycosidase/ ring-opened ribose ΨMP complex was obtained by co-crystallizing ΨMP glycosidase with 4 mM $MnCl_2$, 2 mM R5P and saturated uracil. The K166A/ΨMP complex was prepared by cocrystallizing the mutant enzyme with R5P and uracil.

Data Collection and Processing. Prior to the data collection, crystals were soaked in a cryoprotectant solution of 5% glycerol in the mother liquor to avoid damage during vitrification. Data sets for ΨMP glycosidase and the complexes were collected at the Northeast Collaborative Access Team (NE-CAT) beamline 24-ID-C at the Advanced Photon Source (APS) using an ADSC Quantum 315 detector (Area Detector Systems Corporation) at a wavelength of 0.9795 Å with 1 s exposure times and 1° oscillation angles. The data set for the K166A/ΨMP complex was collected at Cornell High Energy Synchrotron Source (CHESS) using an ADSC Quantum 270 detector at a wavelength of 0.9180 Å with 1 s exposure times and 1° oscillation angles. The detector distances were 275, 350, 325 and 200 cm for the ΨMP glycosidase, the R5P complex, the ring-opened ribose ΨMP complex and the K166A--ΨMP complex, respectively. the HKL2000 program suite (*29*). Data collection statistics are summarized in Table 2.2.

Structure Determination and Refinement. The structure of EcΨMP glycosidase was determined by molecular replacement using MOLREP (*30*), The search model was a monomer of PDB ID 1VKM, which has 39% sequence identity to ΨMP glycosidase, after modification by CHAINSAW (*31*). 1VKM was originally reported to be an indigoidine (IndA)-like protein from *Thermotoga maritima* (*26*), but later shown to be a ΨMP glycosidase (*4*). The model was refined through successive rounds of manual model building using COOT (*32*) and restrained refinement with REFMAC5 (*33*). Water molecules were then included after the model converged, followed by two additional rounds of refinement. The structure of the ΨMP

glycosidase/ring-opened R5P adduct was determined by using the structure of unliganded ΨMP glycosidase as the model and refining with REFMAC5 (*33*). The structure of the ΨMP glycosidase/ ring-opened ribose ΨMP adduct, which crystallized in a different space group, was determined by molecular replacement using MOLREP (*30*) using a monomer from the refined ΨMP glycosidase as the search model. The structure of the K166A/ΨMP complex was determined by using the structure of the ΨMP glycosidase/ ring-opened ribose ΨMP adduct as the model and refining with REFMAC5 (*33*). The ligand contents and alternative side chain conformations were determined by F_o-F_c maps and composite omit maps from PHENIX (*34*). All ligands were clearly observed in the composite omit map, but were not placed until the last stage of the refinement. The data collection and refinement statistics are summarized in Table 2.3.

HLPC Analysis of the Reaction Mixture. HPLC analysis following a linear gradient, at a flow rate of 1 mL/min, was used with absorbance detection at 254 nm. Solvent A was water, solvent B was 100 mM K_2HPO_4, pH 6.6 and solvent C was methanol: 0 min, 100% B; 5 min, 10% A, 90% B; 7 min, 25% A, 60% B, 15% C; 17 min, 25% A, 60% B, 15% C; 19 min, 30% A, 40% B, 30% C; 21 min, 100% B; 30 min 100% B. The column used was a Supelcosil LC-18- T HPLC column (15 cm x 4.6 mm, 3 μM particle size).

Time Course for the Reverse ΨMP Glycosidase Reaction. The reaction mixture contained 20 μM uracil, R5P, and 250 μM $MnCl_2$. The reaction mixture was divided into 90 μL aliquots and 10 μL of 10 μM ΨMP glycosidase was added to achieve a final concentration of 1 μM ΨMP glycosidase. The reaction was quenched at 20 s, 40 s, 1 min, 2 min, and 5 min by heating it at 100 °C for 2 min. The samples were filtered (10 kDa MWCO) and analyzed by RP-HPLC using the method described above. The peak areas for ΨMP were integrated and plotted against time.

Table 2.2. Data collection statistics

	ΨMP glycosidase	ΨMP glycosidase/ ring-opened R5P	ΨMP glycosidase /ring-opened ribose ΨMP	K166A/ΨMP
beamline	APS 24-ID-C	APS 24-ID-C	APS 24-ID-C	CHESS F1
wavelength (Å)	0.9792	0.9795	0.9795	0.9180
space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a (Å)	62.1	61.3	61.5	60.7
b (Å)	115.4	116.5	77.1	76.4
c (Å)	132.0	132.2	200.0	199.0
chains per asymmetric unit	3	3	3	3
resolution (Å)	43.5-2.0 (1.94-1.96)[a]	41.2-2.2 (2.19-2.32)	48.1-2.5 (2.50-2.57)	35.3-1.8 (1.80-1.82)
total no. of reflections	404,995	196,780	124,598	422,749
number of unique reflections	70,995	49,316	33,815	86,433
redundancy	5.7 (5.5)[a]	4.0 (4.0)	3.7 (3.4)	4.9 (4.9)
R_{merge} (%)[b]	5.5 (34.3)[a]	6.0 (40.9)	7.5 (44.1)	7.3(43.1)
completeness (%)	99.7 (99.6)[a]	99.8 (99.9)	99.1 (96.9)	99.1(100)
I/σ(I)	25.1 (4.2)[a]	17.6 (2.4)	17.5 (3.1)	20.4 (3.6)

[a]Values in parentheses are for the highest resolution shell.
[b]$R_{merge} = \Sigma\Sigma_i \mid I_i - \langle I \rangle \mid / \Sigma \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices h,k,l.

Table 2.3. Data refinement statistics

	ΨMP glycosidase	ΨMP glycosidase/ ring-opened R5P	ΨMP glycosidase /ring-opened ribose ΨMP	K166A/ΨMP
no. of reflections in working set	66,972	45,946	31,542	81,618
R_{work}/ R_{free}[a] (%)	18.5/22.0	17.9/23.1	18.5/26.1	17.7/20.7
no. of protein atoms	6,423	6,533	6,399	6,575
no. of ligand atoms	18	69	93	66
no. of water atoms	375	328	210	644
average B-factor protein ($Å^2$)	35.0	41.4	53.5	27.3
average B-factor water ($Å^2$)	34.1	37.3	39.9	31.3
average B-factor ligand ($Å^2$)	57.7	39.5	49.5	17.6
rmsd for bonds (Å)	0.008	0.007	0.008	0.007
rmsd for angles (°)	1.2	1.2	1.3	1.2

[a] $R = \Sigma_{hkl} | |F_{obs}| - k |F_{cal}| | / \Sigma_{hkl} |F_{obs}|$ where F_{obs} and F_{cal} are observed and calculated structure factors. R_{work} = is calculated over all reflections used in the refinement. R_{free} is calculated over a subset of reflections (5%) excluded from all stages of refinement.

For the K166A mutant, the reaction mixture (1 mL) contained 500 μM each of uracil, R5P, and 250 μM of $MnCl_2$. The reaction mixture was divided into 90 μL aliquots and 10 μL of 200 μM K166A was added to achieve the final concentration of K166A (20 μM). The reaction was quenched at 20 s, 40 s, 1 min, 2 min, 3 min, 5 min, 10 min and 20 min time points by heating each reaction mixture at 100 °C for 2 min. The reaction samples were filtered (10 kDa MWCO) and analyzed by RP-HPLC. The peak area for ΨMP was integrated and plotted against time.

Determination of k_{cat} and K_m for the Reverse ΨMP Glycosidase Reaction. Uracil concentrations were varied while keeping the R5P concentration at saturation (1 mM) and the $MnCl_2$ concentration at 250 μM. ΨMP formation was monitored by HPLC analysis. Each reaction mixture was divided into 90 μL aliquots and 10 μL of 10 μM ΨMP glycosidase was added to achieve the final concentration of 1 μM. For each set of reactions, the reaction was quenched at 20 s, 40 s, 1 min, 2 min, 3 min, 5 min, and 10 min time points by heating at 100 °C for 2 min. In the case of K166A, the uracil concentrations were varied while keeping R5P concentration at 1 mM and $MnCl_2$ concentration at 250 μM for five different sets of reactions. Each reaction mixture was divided into 85 μL aliquots and 15 μL of 250 μM K166A was added to start the reaction (final concentration 37.5 μM). For each set of reactions, the reaction was quenched at 20 s, 40 s, 1 min, 2 min, 3 min, 5 min, and 10 min time points by heating at 100 °C for 2 min. All samples were analyzed by HPLC and the amount of ΨMP formed was plotted against time for each set of substrate concentrations. The initial slopes from each set of reactions were then plotted against the substrate concentration and fit to the Michaelis-Menten equation using KaleidaGraph (Synergy Software). The mutant E31A, K93A, H137A, D149A, and N289A activities were measured using a similar procedure.

Biochemical Characterization of the ΨMP Glycosidase R5P Adduct. A reaction mixture containing 2 mM R5P and 1 mM Mn in 50 mM Tris-HCl, 100 mM NaCl and 2 mM TCEP (pH 8.0) was incubated at room temperature for 1 h with ΨMP glycosidase (50 μM), followed by addition of sodium borohydride (2 mg) and further incubation at room temperature for 45 min. The sample was then buffer-exchanged into 10 mM ammonium acetate (pH 8.0) using a Bio-Rad desalting column and analyzed by ICR-MS. Control samples, one containing 2 mM uracil and the other lacking R5P and $MnCl_2$ were similarly prepared and analyzed.

To locate the site of adduct formation, trypsin digestion analysis was performed on each of these samples as follows: 10 μL of guanidine-HCl (6 M) was added to each sample (50 μL containing 100 μg of ΨMP glycosidase), followed by 1 μL of DTT (200 mM). After a 1 h incubation at room temperature, 10 μL of iodoacetamide (200 mM) was added and the reaction mixture was further incubated in the dark for 1 h. 29 μL of ammonium bicarbonate buffer were further added to reduce the final concentration of guanidine-HCl to 0.6 M (final volume 100 μL). 1 μL of trypsin in 50 mM ammonium bicarbonate buffer, pH 8.0 (1 μg/ μL) was added and the samples were incubated at 37 ºC for 20 h. The samples were then analyzed by LC-MS.

Circular Dichroism Measurements. The ΨMP glycosidase protein was in the buffer of 10 mM Tris, 100 mM NaCl, pH 8.0, and used at a final concentration of 7 μM. Data was collected for both the wild-type enzyme and the D149A mutant, either with or without the addition of 300 μM $MnCl_2$. The spectra were collected using an AVIV Biomedical (Lakewood, NJ) Circular Dichroism Spectrometer, Model 202-01. The data were collected at 22 °C from 200 to 250 nm with a 1-nm step size, a 1-nm bandwidth and 1mm-cell length. The spectra represented the average of 10 replica scans. Program K2D2 was used to deconvolute secondary structure contents.

Section 2.3. Results

Overall Structure of ΨMP Glycosidase. ΨMP glycosidase is a homotrimer (Figure 2.2A) with one trimer per asymmetric unit in the crystal structure. Of the 312 possible residues, each protomer is complete except for the first two to five residues at the N-terminus and the C-terminal residue, which is missing in most of the protomers. ΨMP glycosidase shows an αβα fold. The central mixed β-sheet contains 11 strands with topology 1↑4↑3↓2↓7↓8↓11↓9↓10↓5↑6↑ and is flanked by six α-helices on one side and seven α-helices on the other (Figure 2.2 B, C). The structure of the protomer is similar to that of PDB ID 1VKM (*26*), which was used as a search model during molecular replacement. The trimer contains a buried surface area of 8650 $Å^2$, accounting for 27.8 % of the total surface area. The trimer interface is mainly formed by helices α5 and α8 from one protomer and helices α9 and α10 from the adjacent protomer. The interface includes hydrogen bonds between Ile146-Thr180* (* indicates residues from an adjacent protomer) and Arg96-Asn228*, and a salt bridge between Arg97 and Glu179*.

Mn(II) Binding Site. Mn(II) is required for ΨMP glycosidase activity (*4*) and for the formation of high quality crystals. The Mn(II) coordination is octahedral with Mn(II) to oxygen distances ranging from 2.13 Å to 2.35 Å (Figure 2.3). Only one protein residue, Asp145, participates in coordination and water molecules (W1-W5) occupy the remaining five positions. The coordinating oxygen atoms interact with protein atoms, substrate atoms (see below), and other water molecules. W1 hydrogen bonds to Glu179* and a ΨMP phosphate oxygen atom, W2 hydrogen bonds to His137 and a ΨMP phosphate oxygen atom, W3 hydrogen bonds to Asp145, the carbonyl oxygen atom of Gly266, W4 hydrogen bonds to Glu176* and Glu179*, W5 hydrogen bonds to a water molecule in the second sphere (W11) and a water molecule (W7) that

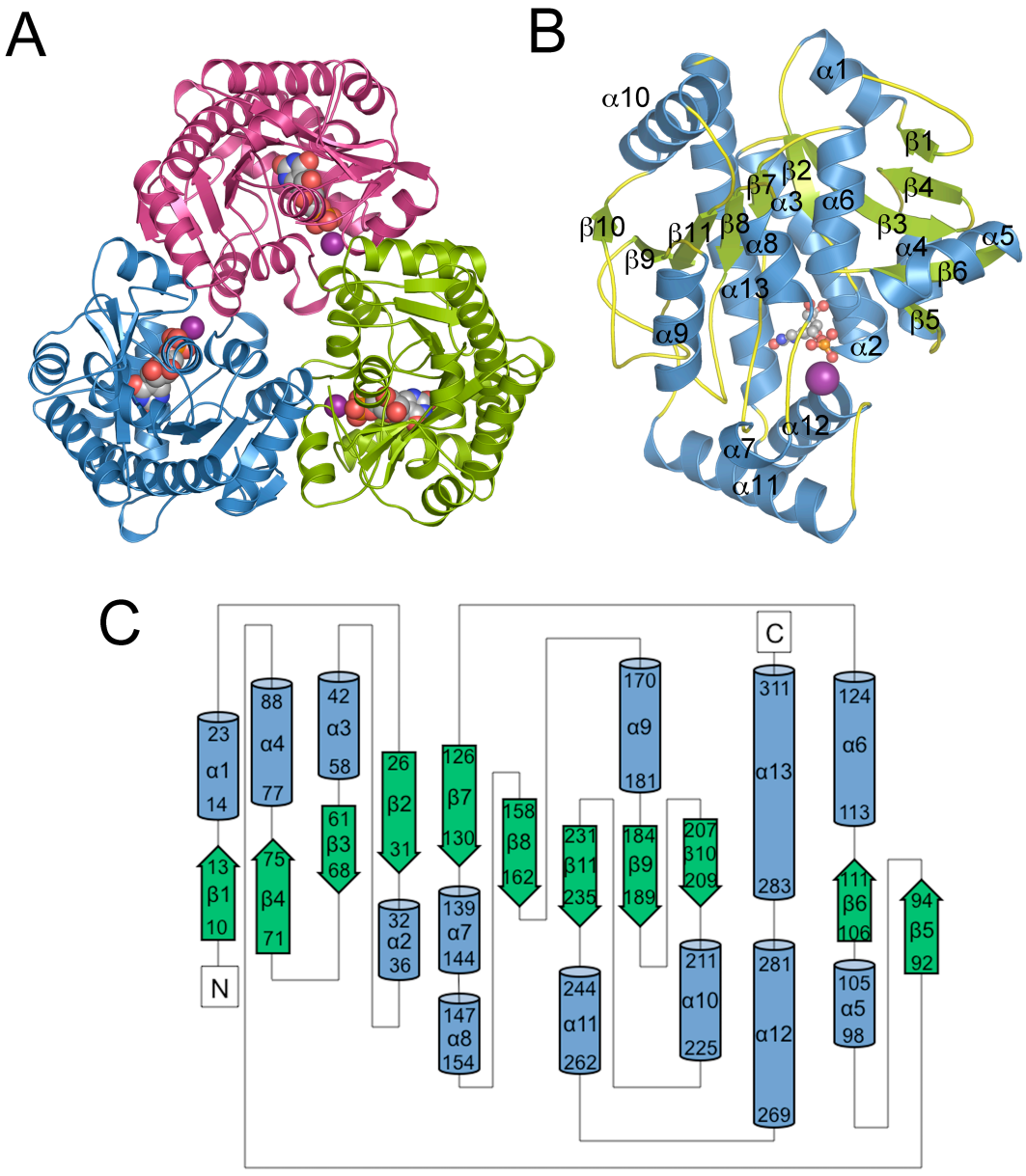


Figure 2.2. Structure of ΨMP glycosidase. (A) Ribbon diagram of the ΨMP glycosidase trimer color coded by protomer. The active site is indicated by the space filling model. The Mn(II) ion is shown in purple. (B) Ribbon diagram of a ΨMP glycosidase protomer labeled with color coded α-helices (blue) and β-strands (β-strands (*35*). (C) Topology diagram of ΨMP glycosidase. The green arrows represent β-strands and blue cylinders α-helices. The first and last residue number of each secondary structural element is indicated.

bridges to the ΨMP phosphate. The water molecules directly involved in Mn(II) interactions are involved in additional hydrogen bonding interactions that extend into the remainder of the active site.

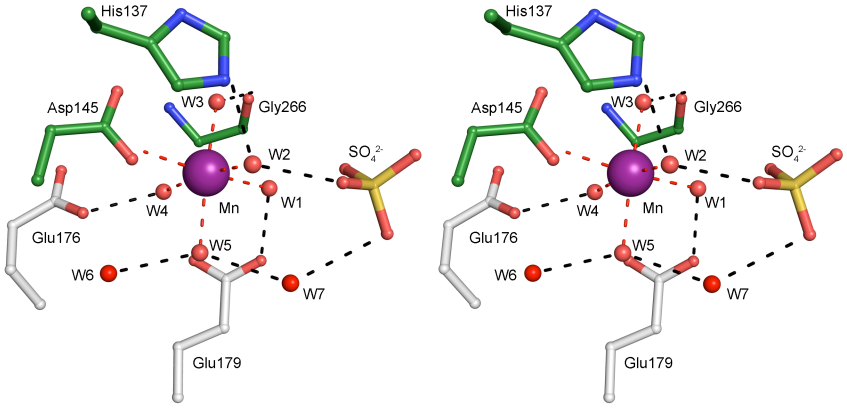


Figure 2.3. Stereoview of Mn(II) binding site. Water molecules and protein atoms in the first and second coordination sphere are shown. The sulfate ion results from the crystallization conditions and occupies the phosphate binding site.

ΨMP Glycosidase Active Site. The active site, as defined by the structure of the K166A/ΨMP complex, is located in a cleft formed by helices α12, α13 and the loops following strands β2, β7, and β8 (Figure 2.4 A, B). The ΨMP ribose is in a C3′-endo conformation, the glycosidic torsion angle is *anti*, and the C4′-C5′ bond is in *gauche, trans* conformation**.** Including the Mn(II) coordination sphere, the active site contains 19 well-ordered water molecules that are present in all three protomers. ΨMP is surrounded by 15 of the water molecules and these mediate many of the active site contacts with the protein. One of the ΨMP phosphate oxygen atoms forms hydrogen bonds with W1 from the Mn(II) coordination sphere, Lys93, and water molecule W11. A second phosphate oxygen atom forms hydrogen bonds with W2 from the Mn(II) coordination sphere, Ser147 and water molecule W10. The third phosphate oxygen atom forms hydrogen bonds with three water molecules (W7-W9), which interact through multiple contacts with the protein (Ser95, Thr112, Ala148NH,

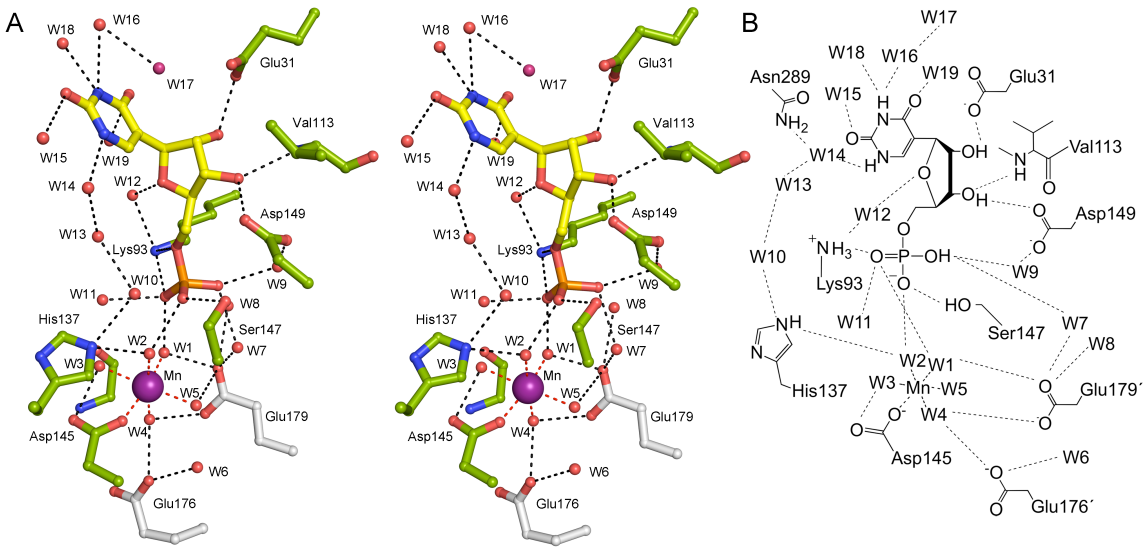


Figure 2.4. ΨMP glycosidase active site. (A) Stereoview of the K166A/ΨMP active site. ΨMP was observed in the active site when cocrystallizing K166A with either ΨMP or R5P and uracil. ΨMP and Mn(II) form extensive hydrogen bond connections with the active site residues and water molecules. The dashed lines represent interactions provided by the residues that make up the hydrogen bond network. (B) Schematic diagram of K166A/ΨMP.

Asp149, and Glu179*). The ribose O2′-hydroxyl group hydrogen bonds to Glu31, the O3′-hydroxyl group hydrogen bonds to Asp149 and the Val113 amide group, while O4′ hydrogen bonds with water molecule W12, which in turn hydrogen bonds to Lys93 and W19.

The ΨMP uracil forms no direct hydrogen bonds with protein residues; however, N1, O2, N3, and O4 form hydrogen bonds with five water molecules. One water molecule (W14) hydrogen bonds to N1 and W15 and bridges to the ΨMP phosphate through two additional water molecules (W10 and W13). The second water molecule (W16) hydrogen bonds to O2, the side chain of Asn289, W17, and W18, which is positioned over the center of the uracil ring and hydrogen bonds to the amide nitrogen and carbonyl oxygen atoms of Gly132. The third water molecule (W15) also hydrogen bonds to O2, the amide nitrogen atom of Ala166 and W14. The fourth water

molecule (W18) hydrogen bonds to N3, the carbonyl oxygen atom of Gly38, and the side chain of Asn289. The fifth water molecule (W19) hydrogen bonds to O4, the carbonyl oxygen atom of His37 and an additional water molecule (W12), which hydrogen bonds to O4′ of ΨMP and Lys93.

Structure of ΨMP Glycosidase/Ring-opened Ribose ΨMP Adduct. The structure of ΨMP glycosidase cocrystallized with R5P and uracil shows ring-opened ribose ΨMP covalently attached to Lys166. Ring opening and closing is correlated with an approximately 90° rotation and a 2 Å shift of the uracil (Figure 2.5 A, B), thus occupying a different binding site compared to the K166A/ΨMP complex. In this binding site, O2 forms a hydrogen bond with Asn289. The ring-opened ribose retains hydrogen bonds between the 2′-hydroxyl group and Glu31, and between the 3′-hydroxyl group and Asp149; however, the hydrogen bond to O3′ requires a rotation of the Glu31 carboxylate about the Cγ-Cδ bond compared to the ΨMP complex. The phosphate group maintains the same position as in the K166A/ΨMP structure and forms hydrogen bonds with Lys93, Ser147, W1, W2, and W7-W11. The water molecules near the uracil are not well defined, possibly owing to the lower resolution (2.5 Å) of this structure.

Structure of the ΨMP Glycosidase/R5P Adduct. Crystals of ΨMP treated with R5P show a structure in which the R5P is in the ring-opened form and is covalently attached to Lys166 via an imine. The R5P superimposes closely with the ring-opened ribose in the ΨMP adduct and forms hydrogen bonds between the 2-hydroxyl group and Glu31, and between the 3-hydroxyl group and Asp149 (Figure 2.5 B, C). The phosphate binding site is essentially the same as for the previous complexes. Water molecules in the uracil binding site are less ordered and more variable from protomer to protomer compared to the K166A/ΨMP complex.

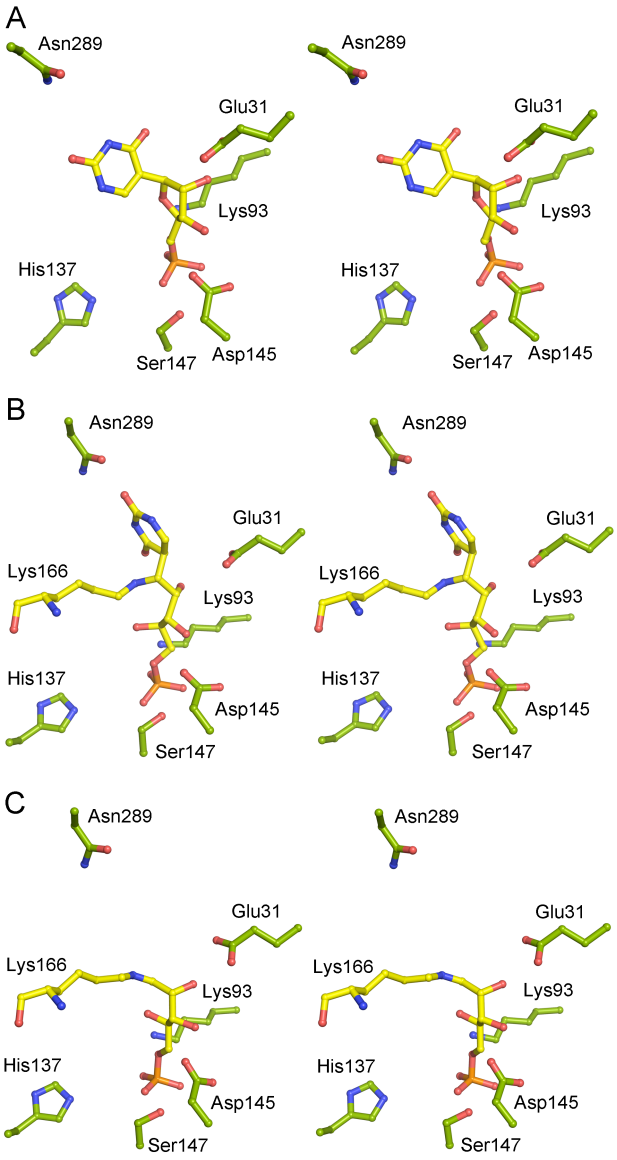
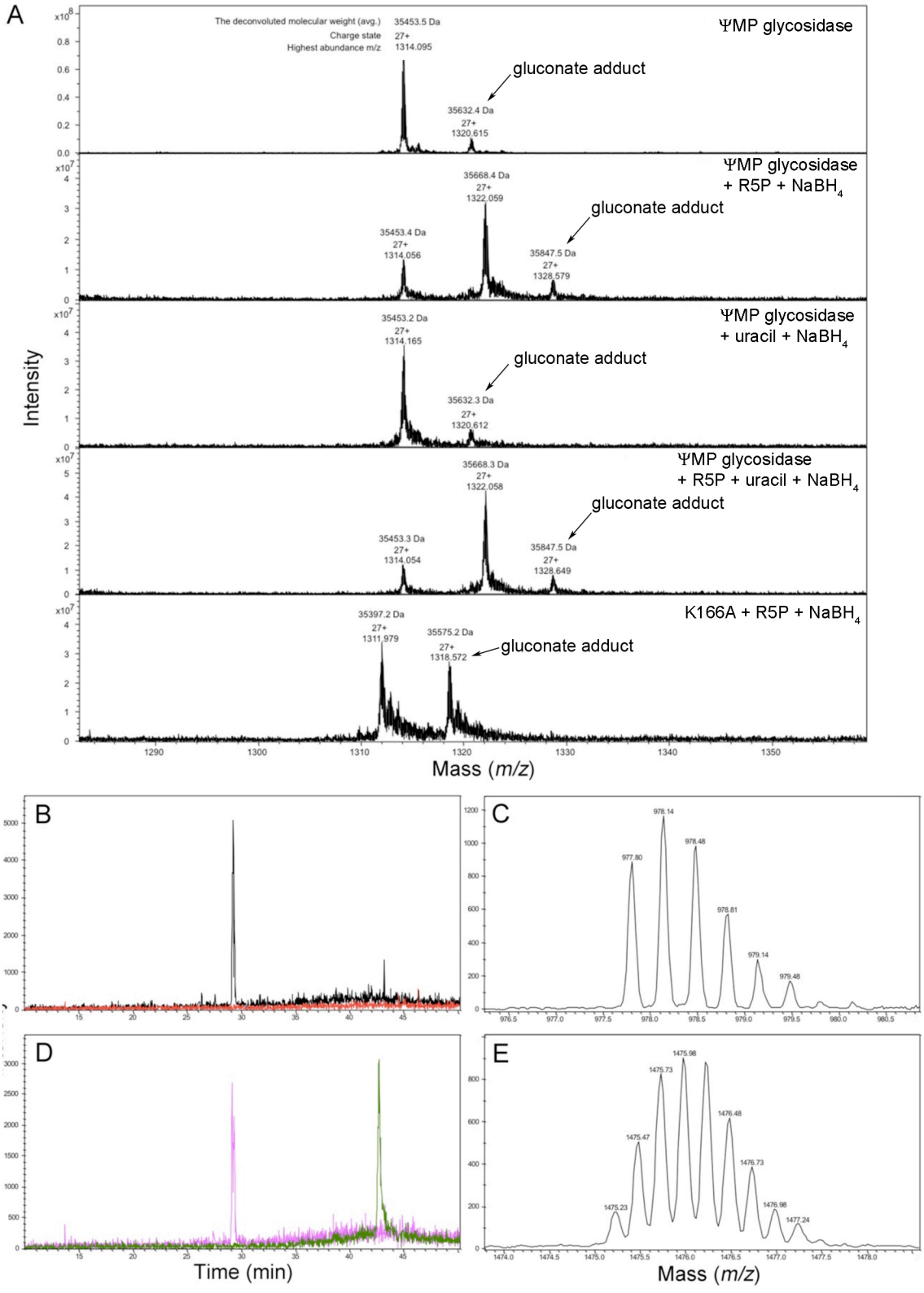


Figure 2.5. Comparison of ΨMP glycosidase complexes. (A) Stereoview of the K166A/ΨMP complex. (B) Stereoview of the ΨMP glycosidase/ ring-opened ribose ΨMP adduct. (C) The ΨMP glycosidase/R5P adduct. Protein carbon atoms are shown in green; ligand carbon atoms in yellow.

Detection of a Lys166 R5P Covalent Adduct by Mass Spectrometry. The mass of wild type ΨMP glycosidase was determined to be 35,453.7 Da by ICR-MS analysis (Figure 2.6 A). The mass of the reduced R5P complex was 35,669.7 Da corresponding to a mass increase of 216 Da. This is consistent with a doubly

Figure 2.6. Analysis of the R5P adduct by mass spectrometry. (A) ICR-MS data for the identification of the borohydride reduced ΨMP glycosidase/R5P adduct. The observed 179 Da adduct corresponds to enzyme gluconoylation (*36*). (B) MS chromatograms of the control sample. The black trace corresponds to the unmodified peptide while the red trace refers to the R5P bound peptide, which is not seen in the ΨMP glycosidase control. (C) Triply charged state of the unmodified peptide in the control ΨMP glycosidase sample. (D) MS chromatograms of the reduced ΨMP glycosidase/R5P adduct. The pink trace corresponds to the unmodified peptide while the green trace corresponds to the R5P bound modified peptide. (E) Triply charged state of the R5P bound modified peptide in the reduced ΨMP glycosidase/R5P adduct.


A
ΨMP glycosidase
The deconvoluted molecular weight (avg.) 35453.5 Da
Charge state 27+
Highest abundance m/z 1314.095
gluconate adduct
35632.4 Da
27+
1320.615
ΨMP glycosidase
+ R5P + NaBH4
35668.4 Da
27+
1322.059
gluconate adduct
35453.4 Da
27+
1314.056
35847.5 Da
27+
1328.579
ΨMP glycosidase
+ uracil + NaBH4
35453.2 Da
27+
1314.165
gluconate adduct
35632.3 Da
27+
1320.612
ΨMP glycosidase
+ R5P + uracil + NaBH4
35668.3 Da
27+
1322.058
gluconate adduct
35453.3 Da
27+
1314.054
35847.5 Da
27+
1328.649
K166A + R5P + NaBH4
35397.2 Da
27+
1311.979
35575.2 Da
27+
1318.572
gluconate adduct
Intensity
Mass (m/z)
B
C
978.14
977.80
978.48
978.81
979.14
979.48
D
E
1475.98
1475.73
1475.47
1476.48
1476.73
1475.23
1476.98
1477.24
Time (min)
Mass (m/z)

protonated reduced R5P-derived imine (216 Da). Adduct formation was not observed when only uracil was incubated with ΨMP glycosidase.

The site of imine formation was identified by trypsin digestion of the reduced ΨMP glycosidase R5P adduct (Figure 2.6 B-E). Trypsin cleaves after lysine and arginine residues and lysine modification blocks the cleavage reaction. Trypsin digestion of the unmodified enzyme yielded the peptide GAEHTFDISADLQELANTNVTVVCAGA**K** (Mass = 2930.4079 Da, amino acids 139-166). For glycosidase incubated with R5P, trypsin digestion yielded the peptide GAEHTFDISADLQELANTNVTVVCAGA**K**SILDLGLTTEYLETFGVPLIGYQTK (Mass = 5896.8759 Da, amino acids 139-191). This demonstrates that Lys166 is the R5P modified residue. This was confirmed by similar MS analysis of the K166A mutant for which no mass increase was observed for the R5P treated sample.

Steady State Kinetics of EcΨMP Glycosidase and its Mutants. Steady state kinetic parameters were determined by measuring the production of ΨMP during the reverse reaction while varying the uracil concentration (Figure 2.7). Michaelis-Menten analysis for the native enzyme indicated a K_m value of 169.6 μM for uracil and a k_{cat} 3.74 s^{-1}, corresponding to a catalytic efficiency, k_{cat}/K_m, of 22 x 10^3 M^{-1} s^{-1}.

Steady state kinetic parameters were also determined for the active site mutants E31A, K93A, K166A and N289A (Table 2.3) (Figure 2.7). The activity of D149A was below the detection limit even at an enzyme concentration as high as 100 μM. K93A has a higher K_m value compared to wild type enzyme. E31A, K93A, K166A and N289A have significantly lower k_{cat} values, with E31A and K166A showing the largest effects.

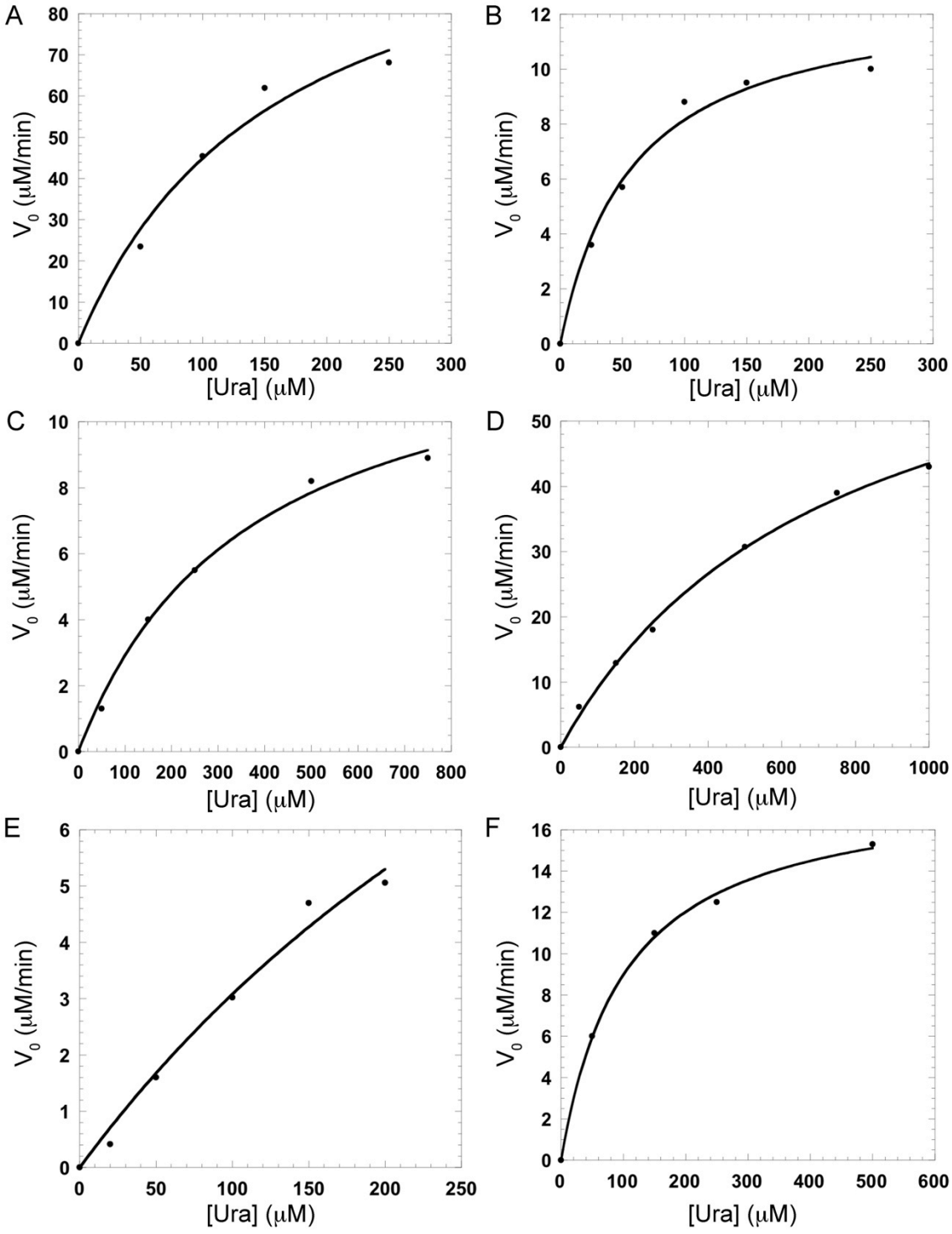


Figure 2.7. Steady state kinetics for ΨMP glycosidase and mutants in the ΨMP formation reaction. (A) ΨMP glycosidase. (B) E31A. (C) K93A. (D) H137A. (E) K166A. (F) N289A.

Table 2.3. Steady state kinetic parameters for ΨMP glycosidase and mutants[a].

Enzyme	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($mM^{-1}\,s^{-1}$)
ΨMP glycosidase	3.74±0.14	171±22	21.9±3.6
E31A	0.0025±0.0001	92.2±11.7	0.027±0.005
K93A	0.22±0.01	295±19	0.75±0.01
H137A	3.22±0.05	589±72	5.44±0.75
D149A	ND[b]	ND	ND
K166A	0.0013±0.000052	154.6±19.4	0.0084±0.0001
N289A	0.224±0.005	190.6±13.8	1.15±0.12

[a] All the kinetic parameters were measured based on the ΨMP formation reaction.
[b] No activity was observed at the level of detection of the assay.

Circular Dichroism for D149A. To determine whether the mutant was properly folded, we examined the CD spectra of the wild-type protein and the D149A mutant. As Mn(II) is shown to be important for the activity of the enzyme, 300 μM $MnCl_2$ was added for comparison. The spectra of the proteins alone and in the presence of Mn(II) are shown in Figure 2.8. The approximate helical content for all the four species are all 84.27% (*35, 37*), indicating very little secondary structural change due to the mutation.

Section 2.4. Discussion

Comparison with Other Protein Structures. A structural similarity search for ΨMP glycosidase was performed using DALI (*38*). Not surprisingly, PDB ID 1VKM, which was used as the search model during molecular replacement, showed the highest similarity with a Z score of 42.2 (sequence identity 39%). PDB ID 1VKM was originally reported by a structural genomics group to be an indigoidine synthase (IndA)-like protein from *T. maritima* (*26*). Subsequent biochemical studies conclusively demonstrated that this protein is a ΨMP glycosidase (*4*). The structure of

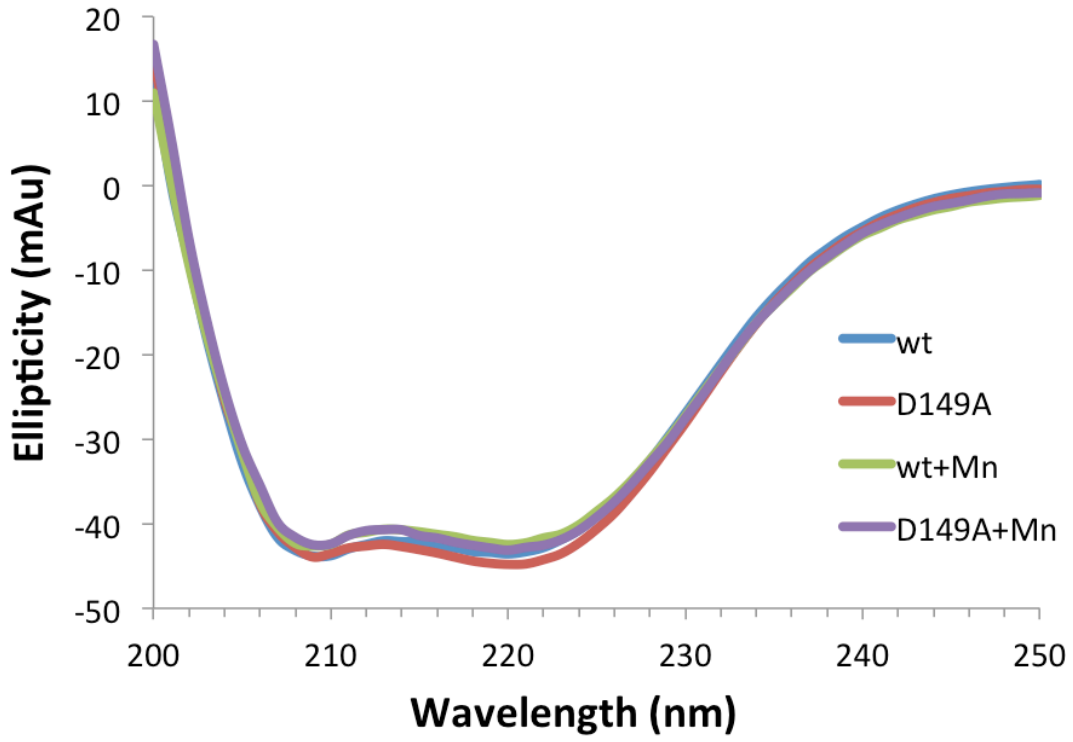


Figure 2.8. The CD spectra of ΨMP glycosidase and the D149A mutant with and without the addition of $MnCl_2$. The CD spectra of the wild type protein was colored blue. The D149A mutant was colored red. The wild type protein with Mn(II) added was colored green. The D149A mutant with Mn(II) added was colored purple.

the EcΨMP glycosidase is trimeric. The structure of ΨMP glycosidase from *T. maritima* (TmΨMP glycosidase) is hexameric in the crystal structure; however, this hexamer is an artifact resulting from face-to-face packing of two trimers, in which pairs of N-terminal His-tags are joined by metal ions. Each of the TmΨMP glycosidase trimers is homologous to the EcΨMP glycosidase trimer, and both enzymes contain a bound Mn(II). The trimer formation is important for the ligand binding since the metal binding site is positioned right between two protomers.

TmΨMP glycosidase was reported to copurify with an unknown bound ligand that resembled a ring-opened sugar phosphate (*26*). Motivated by our structural results, we reexamined this electron density and found that in some protomers a ring-

opened R5P covalently attached to a lysine residue (equivalent to Lys166 in EcΨMP glycosidase) provided a good fit to the electron density while in others the ribose was in the ring-opened form but did not appear to be covalently attached.

Mechanistic Implications. The mechanism of Ψ synthase has been extensively studied and is outlined in Figure 2.9 (*39*). The microscopic reversibility of this reaction would seem to be a reasonable starting hypothesis for the reaction catalyzed by ΨMP glycosidase. In this proposal, tautomerization of ΨMP **2** to give **9** followed by C-glycosyl bond cleavage and trapping of the resulting oxocarbenium ion would give **7** and **8**. Hydrolysis of **8** and protonation of **7** would complete the reaction.

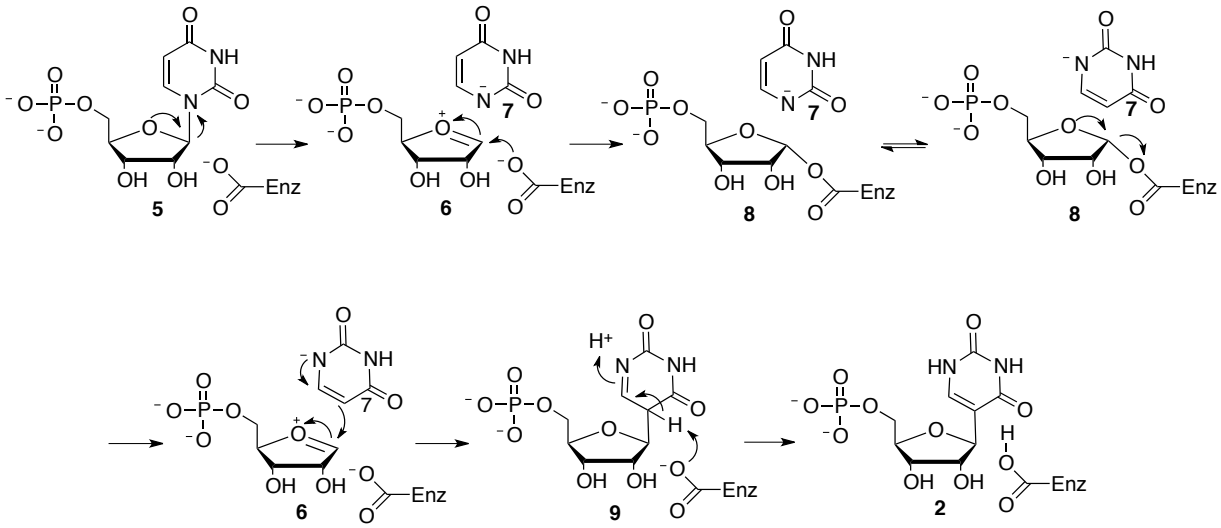


Figure 2.9. Mechanism of pseudouridine formation. This mechanism suggested a plausible starting mechanistic hypothesis for ΨMP glycosidase.

Our structural studies identified four snapshots of the EcΨMP glycosidase reaction coordinate. Wild type EcΨMP glycosidase showed the unliganded state, although a sulfate ion from the crystallization solutions occupied the phosphate binding site. Incubation of the K166A mutant with R5P and uracil resulted in a structure with the substrate ΨMP bound in the active site. Native EcΨMP glycosidase incubated with R5P and uracil showed an intermediate in which ring-opened ribose ΨMP is covalently attached to Lys166. Finally, EcΨMP glycosidase incubated with

R5P alone resulted in a structure with ring-opened R5P attached to Lys166. The last structure is similar to that of the 1VKM complex reported to contain an unknown ligand (*26*). This series of structures demonstrated that the ΨMP glycosidase catalyzed reaction occurs by a mechanism that is significantly different from the microscopic reverse of the Ψ synthase mechanism.

An alternative mechanistic proposal is outlined in Figure 2.10. Ring opening of the ΨMP ribose gives **10**. This reaction requires acid base catalysis. As no amino acids are apparent in the structure for this role (Figure 2.4), the ring opening reaction

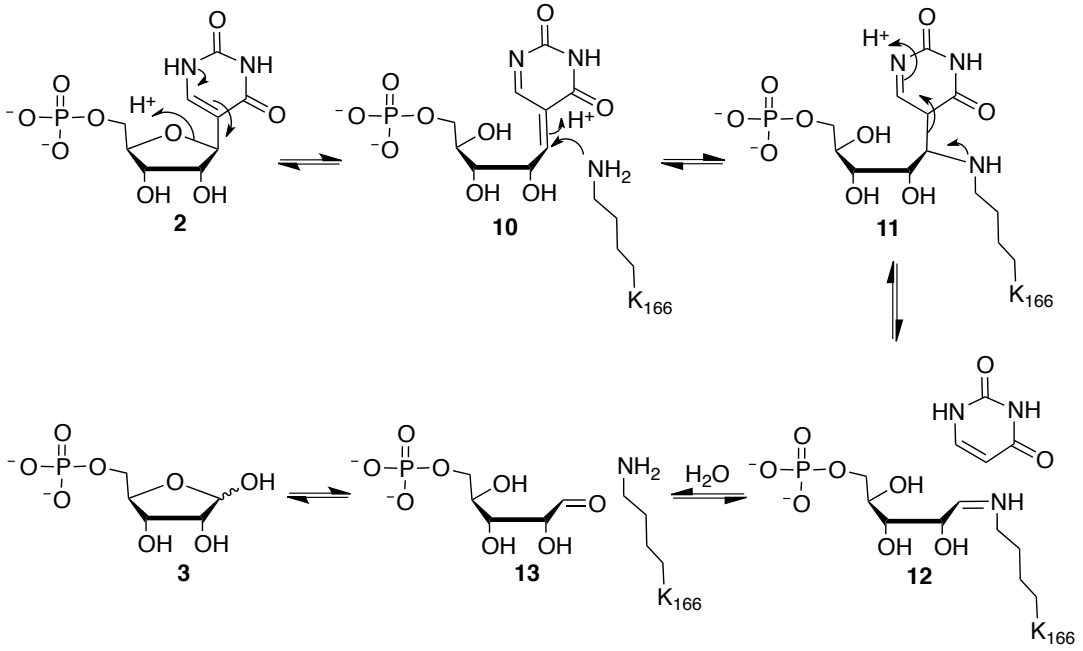


Figure 2.10. Mechanistic proposal for ΨMP glycosidase.

probably occurs by a water-mediated protonation/deprotonation reaction. The conformation of the C-glycosidic bond is reasonable for a concerted process (the C6-C5-C1-O4a dihedral angle is 50º, while the optimal angle would be 90º). Related ribose ring opening reactions are catalyzed by GTP cyclohydrolase (*40*). Following ribose ring opening, Lys166, undergoes a conjugate addition to the C-C double bond of **10** to give **11** with the proton coming from Glu31. Subsequent cleavage of the C-C

glycosidic bond by a retroaldol type reaction releases the uracil anion, which is stabilized by hydrogen bonding of N1 to the hydroxyl of Thr130 and the amide NH of Gly131 and Gly132. Surprisingly, there are no stabilizing hydrogen bonds to the C2 or C4 carbonyl oxygen atoms of uracil. Hydrolysis of imine **12** gives **13**, which will then cyclize to **3**. In addition to the structures, the covalent linkage between Lys166 and the substrate is supported by the characterization of the R5P Lys166 imine by mass spectrometry. An alternative mechanism in which Lys166 displaces the pyrimidine from the ribose, as found for example in 8-oxoguanine DNA glycosylase (*41*), is unlikely. Lysine 166, modeled into the structure of the enzyme ΨMP complex, is not suitably positioned for direct base displacement and the structure of the ring-opened ribose ΨMP complex is not consistent with this mechanism. The structure of UrdGT2, with the substrates modeled into the active site, suggests that this C-C glycosyl transferase does not use an imine intermediate (*24*). Our analysis suggests that ΨMP glycosidase may use a new glycosidase mechanism.

This mechanistic proposal enables us to suggest testable functions for the active site residues. Lys166 plays a key role in facilitating the C-glycosyl bond cleavage reaction and as expected the k_{cat} value of the K166A is 2900-fold lower than wild type, with no change in K_m. It is interesting that low levels of catalytic activity are observed for this mutant, suggesting that water can replace lysine in the mutant. Glu31 is likely to be the proton source for the lysine conjugate addition reaction (**10** to **11** in Figure 2.10). Consistent with this role, the k_{cat} value of the E31A mutant is 7500-fold lower than wild type, with little change in K_m. Lys93 contributes one of several interactions involved in phosphate binding. The k_{cat} value of the K93A mutant is 17-fold lower than wild type, with only a modest increase in K_m. Asn289 hydrogen bonds to two water molecules that are in turn hydrogen bonded to the C2 carbonyl and to N3. These interactions stabilize negative charge on the uracil in the conversion of

11 to **12** and **4**. Consequentially, the k_{cat} value of the N289A mutant is 17-fold lower than wild type, with a small increase in K_m. Asp149 forms a hydrogen bond to the substrate C3′ alcohol. We would therefore expect that the activity of D149A mutant would be comparable to wild type, which is inconsistent with our finding that it shows no detectable activity. One possibility is that the D149A mutant has misfolded; however, circular dichroism spectra of wild type and D149A are nearly identical (Figure 2.8). Another possibility is that Asp149 serves indirectly as an acid/base through a charged network.

The Role of the Hydrated Metal Binding Site. Both EcΨMP glycosidase and TmΨMP glycosidase contain a heavily hydrated, octahedrally coordinated Mn(II) ions in which the ligands are an aspartate side chain and five water molecules. This heavily hydrated metal binding site is unusual and likely plays a role in anchoring the ΨMP phosphate. In addition, the water molecules in the second sphere (W6-W9) are also present in both structures, and all side chains in the second sphere are conserved, including a Glu176 and Glu179 from the neighboring protomer. Two phosphate oxygen atoms form hydrogen bonds to W1 and W2, which are in the Mn(II) coordination sphere. The phosphate group also hydrogen bonds to Ser147 and Lys93, which are conserved in both structures. While the ribose and uracil both undergo movement during catalysis, the phosphate remains in the same position, with no hydrogen bonding changes. ΨMP glycosidase shows highest activities with Fe(II), Co(II) and Mn(II), while Zn(II), and to a lesser extent, Ni(II), were inhibitory (*4*). These observations suggest that the ionic radius of the metal ion may play a role in properly positioning the substrate.

Role of Conformational Changes in Catalysis. Both the protein and substrate undergo conformation changes during catalysis. The uracil undergoes a 90° rotation and a 2 Å shift after ring opening and Lys166 adduct formation. In the glycosidase

ΨMP complex, uracil forms no hydrogen bonds with protein atoms; however, after shifting position, uracil hydrogen bonds through O2 to Asn289, which is located at the opposite end of the active site relative to the phosphate. Repositioning of the uracil also places N1 near the amide nitrogen atoms of Gly131 and Gly132, which are absolutely conserved. Heavy reliance in the binding site on water-mediated interactions is consistent with the need for major repositioning of an intermediate during catalysis.

Comparison of the 12 protomers (three each from four structures) suggests that conformational changes in the protein are also involved in catalysis. The largest differences occur in helices α11 and α12 with RMSDs ranging from 1.5 - 7.0 Å for the pairwise comparisons. At one extreme these residues insert into the active site, acting as a gate. At the other extreme these residues are extended away from the active site making room for substrate binding. In all of the structures, these residues tend to show weaker electron density and higher B-factors. A second region of structural variation occurs in residues 130-150 (α7 and α8) and 162-200 (helix α9 and β9) with RMSDs ranging from 1.5 to 5.0 Å. This region contains 14 of 17 absolutely conserved residues. While crystal packing may also play a role in the variation among protomers, the differences are consistent with gating of the active site and positioning of catalytically important residues.

Conclusion. The ΨMP glycosidase complexes described in this paper give four snapshots of the reaction coordinate and provide key structural insights into the enzymatic hydrolysis of C-glycosides. Most glycosidases utilize a dissociative mechanism to reversibly cleave C-N and C-O glycosidic bonds. In contrast, the structural and biochemical studies reported here demonstrate that ΨMP glycosidase utilizes a mechanism involving ribose ring opening and subsequent covalent linkage

between C1′ and an active site lysine, thus setting the stage for a facile C-glycosyl bond fragmentation by a novel retroaldol type mechanism.

REFERENCES

1. Charette, M., and Gray, M. W. (2000) Pseudouridine in RNA: What, Where, How, and Why, *IUBMB Life 49*, 341-351.
2. Hoang, C., and Ferré-D'Amaré, A. R. (2001) Cocrystal Structure of a tRNA [Psi]55 Pseudouridine Synthase: Nucleotide Flipping by an RNA-Modifying Enzyme, *Cell 107*, 929-939.
3. Uliel, S., Liang, X.-h., Unger, R., and Michaeli, S. (2004) Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organisation, and unique functions, *Int. J. Parasitol. 34*, 445-454.
4. Preumont, A., Snoussi, K., Stroobant, V., Collet, J. F., and Van Schaftingen, E. (2008) Molecular identification of pseudouridine-metabolizing enzymes, *J. Biol. Chem. 283*, 25238-25246.
5. Feng, B., Zheng, M. H., Zheng, Y. F., Lu, A. G., Li, J. W., Wang, M. L., Ma, J. J., Xu, G. W., Liu, B. Y., and Zhu, Z. G. (2005) Normal and modified urinary nucleosides represent novel biomarkers for colorectal cancer diagnosis and surgery monitoring, *J. Gastroen. Hepatol. 20*, 1913-1919.
6. Bililign, T., Griffith, B. R., and Thorson, J. S. (2005) Structure, activity, synthesis and biosynthesis of aryl-C-glycosides, *Nat Prod Rep 22*, 742-760.
7. Gopaul, D. N., Meyer, S. L., Degano, M., Sacchettini, J. C., and Schramm, V. L. (1996) Inosine-uridine nucleoside hydrolase from *Crithidia fasciculata*. Genetic characterization, crystallization, and identification of histidine 241 as a catalytic site residue, *Biochemistry 35*, 5963-5970.

8. Duerre, J. A. (1962) Hydrolytic Nucleosidase Acting on *S*-Adenosylhomocysteine and on 5'-Methylthioadenosine, *J. Biol. Chem. 237*, 3737-&.

9. Hurwitz, J., Heppel, L. A., and Horecker, B. L. (1957) The Enzymatic Cleavage of Adenylic Acid to Adenine and Ribose 5-Phosphate, *J. Biol. Chem. 226*, 525-540.

10. Pugmire, M. J., and Ealick, S. E. (1998) The crystal structure of pyrimidine nucleoside phosphorylase in a closed conformation, *Structure 6*, 1467-1479.

11. Flaks, J. G., Erwin, M. J., and Buchanan, J. M. (1957) Biosynthesis of the Purines .16. the Synthesis of Adenosine 5'-Phosphate and 5-Amino-4-Imidazolecarboxamide Ribotide by a Nucleotide Pyrophosphorylase, *J. Biol. Chem. 228*, 201-213.

12. Olsen, A. S., and Milman, G. (1974) Chinese-Hamster Hypoxanthine-Guanine Phosphoribosyltransferase - Purification, Structural, and Catalytic Properties, *J. Biol. Chem. 249*, 4030-4037.

13. Short, S. A., Armstrong, S. R., Ealick, S. E., and Porter, D. J. T. (1996) Active site amino acids that participate in the catalytic mechanism of nucleoside 2'-deoxyribosyltransferase, *J. Biol. Chem. 271*, 4978-4987.

14. Fromme, J. C., and Verdine, G. L. (2003) Structure of a trapped endonuclease III-DNA covalent intermediate, *EMBO. J. 22*, 3461-3471.

15. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms, *Annu. Rev. Biochem. 77*, 521-555.

16. Kharel, M. K., Pahari, P., Shepherd, M. D., Tibrewal, N., Nybo, S. E., Shaaban, K. A., and Rohr, J. (2012) Angucyclines. Biosynthesis, mode-of-action, new natural products, and synthesis, *Nat. Prod. Rep. 29*, 264-325.

17. Nolan, E. M., Fischbach, M. A., Koglin, A., and Walsh, C. T. (2007) Biosynthetic Tailoring of Microcin E492m: Post-translational Modification Affords an Antibacterial Siderophore-Peptide Conjugate, *J. Am. Chem. Soc. 129*, 14336-14347.

18. Fischbach, M. A., Lin, H., Liu, D. R., and Walsh, C. T. (2005) In vitro characterization of IroB, a pathogen-associated C-glycosyltransferase, *Proc. Natl. Acad. Sci. U.S.A 102*, 571-576.

19. Brazier-Hicks, M., Evans, K. M., Gershater, M. C., Puschmann, H., Steel, P. G., and Edwards, R. (2009) The C-glycosylation of flavonoids in cereals, *J. Biol. Chem. 284*, 17926-17934.

20. Siitonen, V., Claesson, M., Patrikainen, P., Aromaa, M., Maentsaelae, P., Schneider, G., and Metsae-Ketelae, M. (2012) Identification of late-stage glycosylation steps in the biosynthetic pathway of the anthracycline nogalamycin, *ChemBioChem 13*, 120-128.

21. Oja, T., Klika, K. D., Appassamy, L., Sinkkonen, J., Mantsala, P., Niemi, J., and Metsa-Ketela, M. (2012) Biosynthetic pathway toward carbohydrate-like moieties of alnumycins contains unusual steps for C-C bond formation and cleavage, *Proc. Natl. Acad. Sci. U.S.A 109*, 6024-6029, S6024/6021-S6024/6019.

22. Dumitru, R. V., and Ragsdale, S. W. (2004) Mechanism of 4-(β-D-Ribofuranosyl)aminobenzene 5'-Phosphate Synthase, a Key Enzyme in the Methanopterin Biosynthetic Pathway, *J. Biol. Chem. 279*, 39389-39395.

23. White, R. H. (2011) The conversion of a phenol to an aniline occurs in the biochemical formation of the 1-(4-aminophenyl)-1-deoxy-D-ribitol moiety in methanopterin, *Biochemistry 50*, 6041-6052.

24. Mittler, M., Bechthold, A., and Schulz, G. E. (2007) Structure and Action of the C-C Bond-forming Glycosyltransferase UrdGT2 Involved in the Biosynthesis of the Antibiotic Urdamycin, *J. Mol. Biol. 372*, 67-76.

25. Chang, A., Singh, S., Phillips, G. N., Jr., and Thorson, J. S. (2011) Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation, *Curr. Opin. Biotechol. 22*, 800-808.

26. Levin, I., Miller, M. D., Schwarzenbacher, R., McMullan, D., Abdubek, P., Ambing, E., Biorac, T., Cambell, J., Canaves, J. M., Chiu, H. J., Deacon, A. M., DiDonato, M., Elsliger, M. A., Godzik, A., Grittini, C., Grzechnik, S. K., Hale, J., Hampton, E., Han, G. W., Haugen, J., Hornsby, M., Jaroszewski, L., Karlak, C., Klock, H. E., Koesema, E., Kreusch, A., Kuhn, P., Lesley, S. A., Morse, A., Moy, K., Nigoghossian, E., Ouyang, J., Page, R., Quijano, K., Reyes, R., Robb, A., Sims, E., Spraggon, G., Stevens, R. C., van den Bedem, H., Velasquez, J., Vincent, J., Wang, X., West, B., Wolf, G., Xu, Q., Zagnitko, O., Hodgson, K. O., Wooley, J., and Wilson, I. A. (2005) Crystal structure of an indigoidine synthase A (IndA)-like protein (TM1464) from *Thermotoga maritima* at 1.90 Å resolution reveals a new fold, *Proteins. 59*, 864-868.

27. Ausubel, F. M., and Brent, F. (1987) *Curr. Protoc. Mol. Biol.*, John Wiley and Sons, New York.

28. Vita, A., Huang, C. Y., and Magni, G. (1983) Uridine phosphorylase from Escherichia coli B.: Kinetic studies on the mechanism of catalysis, *Arch. Biochem. Biophys. 226*, 687-692.

29. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, In *Macro. Crystallogr., Pt A*, pp 307-326.

30. Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement, *Acta Crystallogr. Sect. D-Biol. Crystallogr. 56*, 1622-1624.

31. Stein, N. (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement, *J. Appl. Cryst. 41*, 641-643.

32. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D 60*, 2126-2132.

33. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S., and Dodson, E. J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT, *Acta Crystallogr. Sect. D-Biol. Crystallogr. 55*, 247-255.

34. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination, *Acta Crystallogr. Sect. D-Biol. Crystallogr. 58*, 1948-1954.

35. Greenfield, N. J., and Fasman, G. D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation, *Biochemistry 8*, 4108-4116.

36. Geoghegan, K. F., Dixon, H. B., Rosner, P. J., Hoth, L. R., Lanzetti, A. J., Borzilleri, K. A., Marr, E. S., Pezzullo, L. H., Martin, L. B., LeMotte, P. K., McColl, A. S., Kamath, A. V., and Stroh, J. G. (1999) Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins, *Anal. Biochem. 267*, 169-184.

37. Deleage, G., and Geourjon, C. (1993) AN INTERACTIVE GRAPHIC PROGRAM FOR CALCULATING THE SECONDARY STRUCTURE-CONTENT OF PROTEINS FROM CIRCULAR-DICHROISM SPECTRUM, *Comput. Appl. Biosci. 9*, 197-199.

38. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res. 38 Suppl*, W545-549.

39. Miracco, E. J., and Mueller, E. G. (2011) The products of 5-fluorouridine by the action of the pseudouridine synthase TruB disfavor one mechanism and suggest another, *J. Am. Chem. Soc. 133*, 11826-11829.

40. Rebelo, J., Auerbach, G., Bader, G., Bracher, A., Nar, H., Hosl, C., Schramek, N., Kaiser, J., Bacher, A., Huber, R., and Fischer, M. (2003) Biosynthesis of Pteridines. Reaction Mechanism of GTP Cyclohydrolase I, *J. Mol. Biol. 326*, 503-516.

41. Chung, S. J., and Verdine, G. L. (2004) Structures of End Products Resulting from Lesion Processing by a DNA Glycosylase/Lyase, *Chem. & Biol. 11*, 1643-1649.

CHAPTER 3

THIAMIN PYRIMIDINE BIOSYNTHESIS IN *CANDIDA ALBICANS*: A REMARKABLE REACTION BETWEEN HISTIDINE AND PYRIDOXAL PHOSPHATE[1]

Section 3.1. Introduction

The thiazole **7** and the pyrimidine **10** heterocycles of thiamin are biosynthesized in Saccharomyces cerevisiae using chemistry that is fundamentally different from the bacterial pathway (1, 2) (Figure 3.1). In bacteria the thiazole is formed from deoxy-D-xylulose 5-phosphate, glycine (or tyrosine) and cysteine and the pyrimidine is formed from aminoimidazole ribotide(3). In contrast, in *S. cerevisiae* the thiazole **7** is formed from NAD, glycine and an active site cysteine of the THI4 protein (4) and the 4-amino-2-methyl-5-hydroxymethyl pyrimidine phosphate (HMP-P) (**10**) is formed from histidine and pyridoxal phosphate (PLP) (5-9). The origin of all of the atoms of HMP-P has been determined as shown in Figure 3.1 and suggests that HMP-P (**10**) is formed using remarkable chemistry that is without chemical or biochemical precedent. Genetic studies have demonstrated that HMP-P formation requires only a single gene (THI5) (10). Here we report the overexpression of THI5p, the reconstitution of the enzymatic activity, the preliminary characterization of the reaction and a structure of the enzyme/PLP complex. These experiments suggest that the THI5 protein is the histidine source for HMP-P formation and that THI5p is a single turnover enzyme.

---

[1] Produced with permission from Lai R., Huang S., Fenwick M. K., Hazra A., Zhang Y., Rajashanka K., Philmus B., Kinsland C., Sanders J. M., Ealick S. E., Begley T. P., (2012), *J. Am. Chem. Soc., 134* (22), 9157-9159.

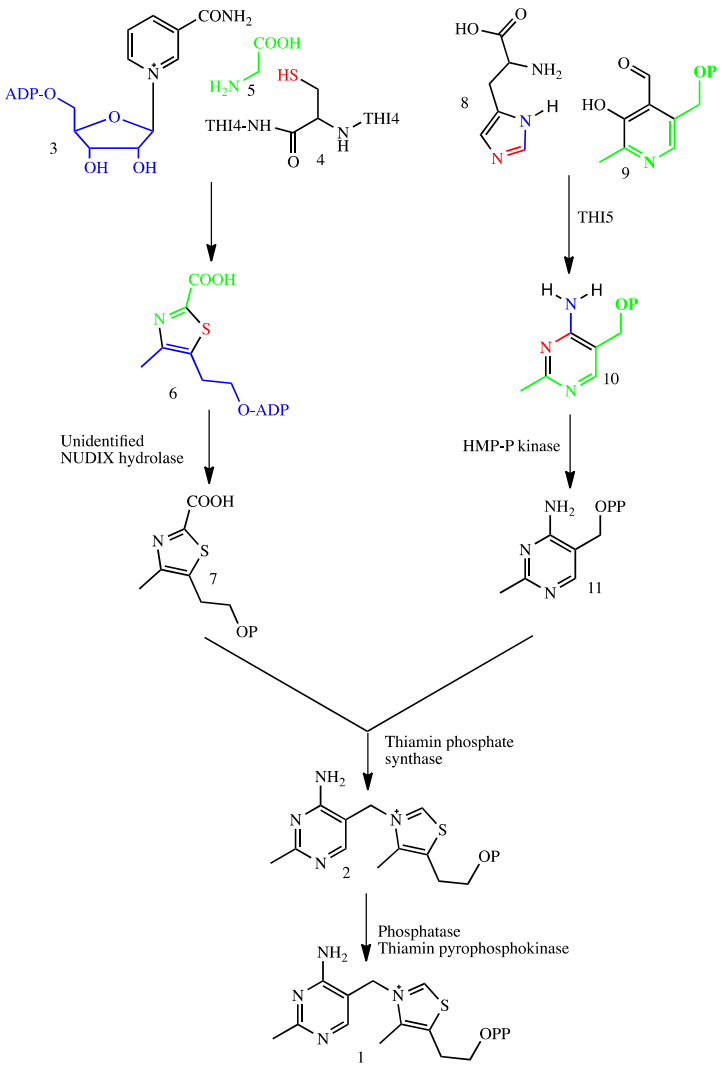


Figure 3.1. Thiamin pyrophosphate biosynthesis in *S. cerevisiae*

Section 3.2. Materials and Methods

THI5p Overexpression and Purification. *E. coli* BL21(DE3) containing the *THI5* gene in pET28b was grown in minimal medium (11.3 g M9 salts, 13.3 mL 50% glucose, 2.7 mL 1M $MgSO_4$, 100 µL 1M $CaCl_2$ diluted to 1 L medium) containing Kanamycin (40 µg/mL) with shaking at 37 °C until the OD_{600} reached 0.6. At this point, protein overexpression was induced with IPTG (final concentration, 500 µM), and cell growth was continued at 15 °C for 16 h. The cells were harvested by

centrifugation, and the cell culture were resuspended in 20 mL of lysis buffer (10 mM imidazole, 300 mM NaCl,pellets from 1 L of 50 mM NaH_2PO_4, 5 mM DTT, pH 8) and lysed by sonication (Heat System Ultrasonics model W-385 sonicator, 1.5 s cycle, 60% duty). The resulting cell lysate was clarified by centrifugation and THI5p was purified on a Ni-NTA column following the manufacturer's instructions. After elution, the protein was desalted under anaerobic conditions using a 10-DG column (BioRad) pre-equilibrated with 100 mM Tris-HCl buffer, 5 mM DTT, 30% glycerol, pH 7.5. The purified protein was stored in aliquots in liquid nitrogen.

Reconstitution of the THI5p Activity. 100 μL of 600 uM THI5p was anaerobically pre-incubated with 1.5 equivalents of $Fe(NH_4)_2(SO_4)_2$ in an ice bath for 30 min. The mixture was anaerobically desalted using a Bio-Spin 6 column (Bio-Rad) pre-equilibrated with 100 mM Tris-HCl, 5 mM DTT, pH 7.5. Subsequently, 2 equivalents of PLP were added. The reaction mixture was aerobically incubated at room temperature for 3 h, filtered using a 10 kDa MW cutoff filter, and analyzed by HPLC and LC-MS. Aerobic preincubation gave lower yields of product (>2 fold reduction). The reconstitution was run using 0.5, 1, 2, 3 and 4 equivalents of PLP. Two equivalents of PLP gave the highest level of product production.

HPLC Conditions for Analyzing the THI5p Reaction mixture. The following linear gradient, at a flow rate of 2 mL/min, on a Supelcosil LC-18 column (250 mm x 10 mm, 5 μm ID) was used: solvent A is water, solvent B is 100 mM potassium phosphate, pH 6.6, solvent C is methanol; 0 min: 100% B; 5 min: 10% A, 90% B; 10 min: 25% A, 60% B, 15% C; 13 min: 25% A, 60% B, 15% C; 17 min: 30% A, 10% B, 60% C; 18 min: 30% A, 10% B, 60% C; 21 min: 100% B; 30min: 100% B.

^{15}N-THI5p Overexpression and Purification. The overexpression and purification procedures were the same as described above for THI5p except $^{14}NH_4Cl$ was replaced by $^{15}NH_4Cl$ in the M9 salts.

LC-MS Analysis of the THI5p reaction mixture. The following linear gradient, at a flow rate of 0.4 mL/min, on a Supelcosil LC-18-T column (150 mm X 3 mm, 3 μm ID) was used: solvent A is 20 mM NH_4OAc, pH 6.6, solvent B is 75% methanol; 0 min: 100% A; 3 min: 100% A; 8 min: 50% A, 50% B; 10 min: 100% A; 20 min: 100% A. The HPLC is coupled with a microTOF-Q II for MS analysis in the positive mode.

Trypsin Digestion of inactive THI5p. After completion of PLP formation, the inactive THI5p (50 μg, 1.8 μl of 660 μM THI5p) was added to 10 μL of 6 M guanidine-HCl, 25 mM ammonium bicarbonate, pH 8. 1 μL of 200 mM DTT in 25 mM ammonium bicarbonate pH 8, was then added to the denatured protein solution and the resulting mixture was incubated at room temperature for 1 h. After that, 10 μL of 200 mM iodoacetamide in 25 mM ammonium bicarbonate pH 8, was added and the reaction mixture was incubated at room temperature in the dark for 1 h. Finally, 77.5 μL of 25 mM ammonium bicarbonate pH 8, was added to reduce the guanidine-HCl concentration to 0.6 M followed by the addition of 1 μg of trypsin. The mixture was incubated at 37 °C overnight and stored at -20 °C until analysis.

LC-MS Conditions for Analyzing THI5p Samples after Trypsin Digestion. The following linear gradient, at a flow rate of 0.4 mL/min, on a Synergi Polar-RP 100A column (50 mm X 2 mm, 2.5 microns) was used: solvent A is 0.1% formic acid in water, solvent B is 0.1% formic acid in acetonitrile; 0 min: 95% A, 5% B; 1 min: 100% A; 60 min: 35% A, 65% B; 61 min: 100% B; 62 min: 100% B; 64 min: 95% A, 5% B; 65.5 min: 95% A, 5% B. The HPLC is coupled with a microTOF-Q II for MS analysis in the positive mode.

THI5p Refolding Experiments. 1 mL of 700 μM THI5p was added to 25 mL of denaturing buffer (50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, 8 M urea, pH 7.8) and incubated at room temperature for 1 h. The solution was then transferred

to a Slide-A-Lyzer Dialysis Cassette, 10K MWCO (Thermo Scientific) and dialyzed against 2 L of dialysis buffer (50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, 5 M urea, pH 7.8) for 5 h. Subsequently, the cassette was transferred to the next buffer (50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, 2M urea, pH 7.8) and dialyzed overnight. The protein solution was then loaded onto a 5 mL Ni-NTA column pre-equilibrated with 50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, 2 M urea, pH 7.8. The Ni-NTA column was installed on an FPLC and eluted, at a flow rate of 1 mL/min, with the following linear gradient: solvent A is 50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, 2 M urea, pH 7.8; solvent B is 50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, pH 7.8; 0 min: 100% A; 120 min: 100% B. Finally, the refolded protein was eluted with 200 mM imidazole, 50 mM KH_2PO_4, 300 mM NaCl, 5 mM mercaptoethanol, pH 7.8. After concentration, the protein solution was buffer-exchanged using a Bio-Spin 6 Column (Bio-Rad) into 100 mM Tris-HCl, 5 mM DTT, pH 7.5. The activity of the refolded THI5p protein was determined as described above for the native enzyme.

Mutagenesis of THI5p. The mutants shown in Table S1 were constructed. The activity of each mutant was determined as described above for the native enzyme. Most of the H18A mutant overexpressed as inclusion bodies, however a very small amount of this mutant could be purified from the crude lysate and was found to be active.

Crystallization of CaTHI5p. Initial attempts to crystallized *S. cerevisiae* THI5p were unsuccessful. We then cloned and overexpressed a collection of THI5p orthologs. Of these, CaTHI5p produced high quality crystals, which were used for all crystallographic studies. For crystallization experiments, the poly-histidine tag of CaTHI5p was removed by thrombin cleavage. A total of 0.45 mg thrombin from bovine plasma (Sigma Aldrich) was added to 8 mL of reaction mixture containing 2.5

mg/mL CaTHI5, 10 mM Tris, pH 7.5, 20 mM NaCl and 1 mM DTT. This solution was incubated overnight at 4 °C with shaking. The sample was subjected to a second round of nickel-chelate chromatography to remove uncleaved protein, followed by size exclusion chromatography using a HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare). The protein purity as judged by SDS-PAGE analysis was higher than 95 %. The protein was buffer-exchanged into 20 mM NaCl, 10 mM Tris, pH 7.5, and concentrated to 15 mg/mL for crystallization experiments.

CaTHI5p was cocrystallized with 1 mM PLP using the hanging-drop vapor diffusion method at 22 °C. The initial drops consisted of a 1 μL of protein solution and 1 μL of reservoir solution. The optimized reservoir solution was 20 % (w/v) PEG 3350 and 0.2 M ammonium chloride. Rod-shaped crystals grew within one day to the size of 500 μm × 10 μm × 10 μm. The crystals were briefly soaked in 20 % (w/v) PEG 3350, 0.2 M ammonium chloride and 15 % glycerol prior to vitrification in liquid nitrogen.

Crystallization of CaTHI5p-H66G. The poly-histidine tag of CaTHI5p-H66G was cleaved following a similar procedure as described above for the wild type protein, except 1 mg thrombin was added to 14 mL of reaction mixture containing 4.5 mg/mL CaTHI5p-H66G, 25 mM Tris and 150 mM NaCl, pH 7.8. The final sample was in a buffer of 30 mM NaCl and 10 mM citrate, pH 6.5, at a protein concentration of ~10 mg/mL for crystallization experiments.

CaTHI5p-H66G was crystallized using the hanging drop vapor diffusion method at 22 °C. The drops consisted of 1.5 μL of protein solution and 1.5 μL of reservoir solution. The optimized reservoir solution contained 15 % (w/v) polyethylene glycol (*11*) 4000 and 100 mM citrate, pH 5.3. The crystals were briefly soaked in a cryosolution containing 25 mM citric acid (pH 5.2), 38 % (w/v) PEG 4000, and 5 mM DTT prior to vitrification in liquid nitrogen.

Structure Determination of CaTHI5p-H66G. The structure of CaTHI5p-H66G was determined by single wavelength anomalous diffraction phasing using CaTHI5p-H66G crystals soaked in the cryosolution supplemented with 1.0 M NaBr for approximately 1 min (*12*). X-ray data were measured at 100 K at NE-CAT beamline 24-ID-C of the Advanced Photon Source. 420 images were recorded from one crystal using a PILATUS-6MF (Dectris) detector. Diffraction data were integrated and scaled to 2.0 Å resolution using HKL2000 (*13*). The space group of the crystal was $P2_12_12_1$ with unit cell parameters a = 55.0 Å, b = 99.8 Å, and c = 126.5 Å. Heavy atom sites were identified using SHELXD (*14, 15*) via HKL2MAP (*16*). SAD phasing, density modification, and initial model building were performed using MLPHARE (*17*), Parrot (*18*), and RESOLVE (*19*), respectively. The model was completed using COOT (*20*) to perform manual building alternated with refinement using REFMAC5 (*21*) and PHENIX (*9*) The final structure of CaTHI5p-H66G, with one homodimer per asymmetric unit, was subsequently refined against a higher resolution data set at 1.6 Å resolution. Complete main chains were built for each protomer and only eight side chains were missing atoms and incomplete. A difference Fourier map showed that each protomer contained a citrate molecule in the proposed active site. The final structure had an R factor of 17.3 % and R_{free} of 20.2 %. X-ray diffraction and refinement statistics are provided in Tables 1 and 2, respectively.

Structure Determination of CaTHI5p/PLP. X-ray diffraction data for CaTHI5p/PLP were collected at 100 K using the Advanced Photon Source NE-CAT beamline 24-ID-C and a PILATUS-6MF detector (Dectris). Crystals of the CaTHI5/PLP complex diffracted to 2.2 Å resolution. The space group was $P2_12_12_1$ and the unit cell parameters (a = 55.16 Å, b = 100.46 Å, and c = 126.20 Å) were similar to those of CaTHI5p-H66G. The data were indexed, integrated and scaled using HKL2000 (*13*). The structure was determined using the structure of CaTHI5-

Table 3.1. Data Collection Statistics.

	Br-CaTHI5p-H66G	CaTHI5p-H66G	CaTHI5p/PLP
beamline	APS 24-ID-C	APS 24-ID-C	APS 24-ID-C
wavelength (Å)	0.9194	0.9792	0.9795
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a (Å)	55.0	55.4	55.0
b (Å)	99.8	99.5	99.7
c (Å)	126.5	126.4	123.4
Resolution range (Å)	50.0–2.0 (2.07–2.00)[a]	50.0–1.6 (1.63–1.60)	48.4–2.2 (2.34–2.20)
No. of reflections	718,132	395,599	182,558
No. of unique reflections	90,983; 43,071[b]	92,106	36,139
Redundancy	7.9 (7.8)	4.3 (4.3)	5.1 (5.0)
Completeness (%)	99.9 (99.9)	99.1 (99.2)	98.9 (99.8)
$\langle I \rangle / \langle \sigma_I \rangle$[c]	18.8 (4.2)	22.7 (3.3)	14.5 (2.3)
R_{merge} (%)[d]	10.4 (56.3)	6.0 (48.1)	10.5 (69.0)

[a]Values in parentheses correspond to the highest resolution shell. [b]Number of Bijvoet pairs. [c]$\langle\ \rangle$ indicates mean; I denotes intensity, and σ_I represents the standard deviation of I. [d] $R_{merge} = \sum_{i=1}^{N_U}\sum_{j=1}^{N_i}\left|I_j - \langle I \rangle_i\right| / \sum_{i=1}^{N_U}\sum_{j=1}^{N_i} I_j$, where N_U is the total number of unique reflections and N_i is the number of reflections measured for a given $i = hkl$.

Table 3.2. Data Refinement Statistics.

	CaTHI5p-H66G	CaTHI5p/PLP
No. of reflections	92,025	36,004
No. of reflections in working set	87,408	34,157
Resolution (Å)	1.60	2.20
No. of protein atoms	5,662	4929
No. of waters	818	318
No. of ligand atoms	26 (citrate)	46 (PLP)
RMSD from ideal bonds (Å)	0.006	0.007
RMSD from ideal angles (*22*)	1.055	1.061
R_{work} (%)[a]	16.2	17.4
R_{free}(%)[a]	19.1	22.8
Ramachandran analysis[b]		
Most favored (%)	98.2	92.7
Additional allowed (%)	1.8	7.3
Generously allowed (%)	0.0	0.0
Disallowed (%)	0.0	0.0

[a] $R_{work} = \sum_{i=1}^{N_U} \left| |F_{O,i}| - k|F_{C,i}| \right| / \sum_{i=1}^{N_U} |F_{O,i}|$, where $|F_{O,i}|$ and $|F_{C,i}|$ are the observed and calculated structure factor amplitudes for reflection $i = hkl$ and N_U is the number of unique reflections. For R_{free}, these sums were taken over a 5% subset of the reflections excluded during structure refinement. [b]Carried out using PROCHECK (excludes glycine, proline, and end residues) (*23*).

H66G as the starting model and refined using REFMAC5 (*21*) and PHENIX (*9*). Water molecules and one PLP molecule per protomer were added after the initial refinement converged. The final structure had an R factor of 21.3 % and R_{free} of 23.5 %. Data collection and refinement statistics are provided in Tables 3.1 and 3.2, respectively.

Section 3.3. Results

Reconstitution of THI5p Activity. The THI5 gene from C. albicans was overexpressed in E. coli BL21(DE3), grown in minimal medium from a pET28b vector and purified by Ni-NTA chromatography. The purified protein, has the expected molecular mass of 40408 Da. The active form of the enzyme was prepared by treating this protein with Fe(II) or Fe(III) under anaerobic conditions followed by removal of excess iron. Addition of PLP to this preparation, under aerobic conditions, yielded HMP-P (Figure 3.2). This reaction product was characterized by comigration with an authentic standard, by dephosphorylation to give HMP, by enzymatic conversion to

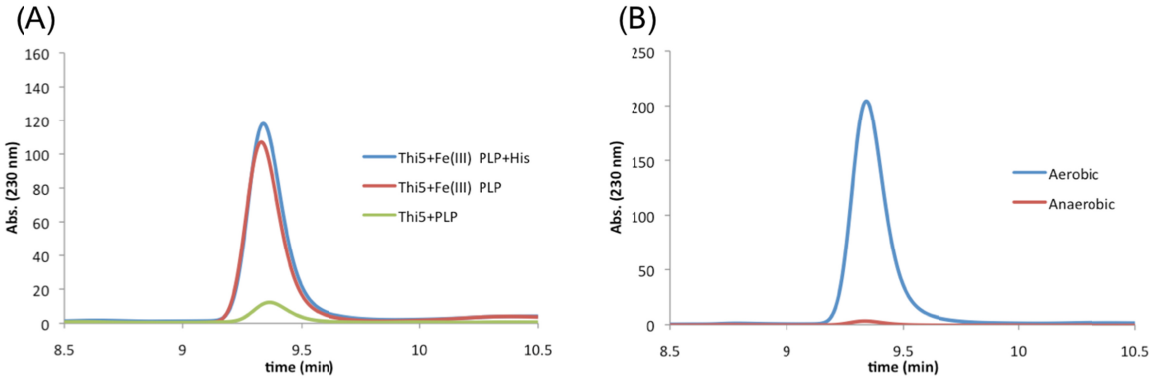


Figure 3.2. Reconstitution of the THI5 catalyzed reaction. (A) The reconstitution reaction requires PLP and Fe(III) but is independent of added histidine. The small amount of HMP-P shown in the green trace is due to product that copurifies with THI5 and not to synthesis. (B) The HMP-P forming reaction is oxygen requiring.

thiamin monophosphate and by MS analysis. After optimization and scale-up, it was also possible to characterize the product by ^{1}H NMR analysis. Two remarkable features of this reaction are: 1) HMP-P formation does not require the addition of histidine or a histidine-derived metabolite and 2) the reaction was not catalytic (THI5p:HMP-P = 1:0.5).

Labeling analysis of THI5p. ^{15}N-THI5p, prepared by growing the overexpression strain in ^{15}N-ammonium chloride-containing minimal medium, produced HMP-P that upon MS analysis showed the expected M+2 peak corresponding to incorporation of 2 atoms of ^{15}N from the protein (Figure 3.3). This analysis of the labeled and unlabeled proteins by ESI-MS also confirmed a high level of ^{15}N incorporation in the ^{15}N-THI5p (Figure 3.4).

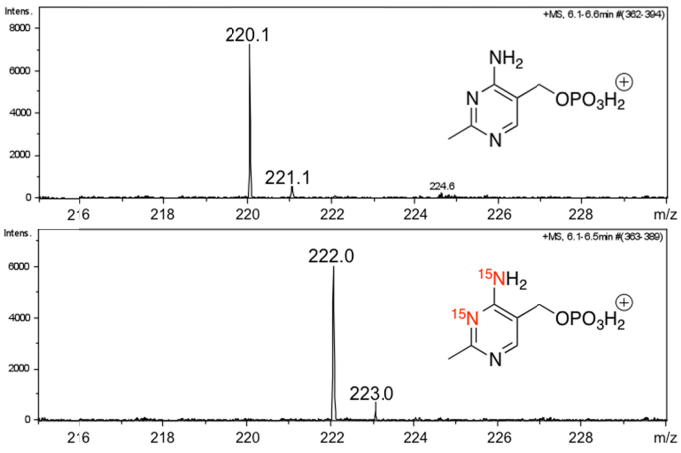


Figure 3.3. MS analysis of HMP-P formed using A) ^{14}N-THI5p and B) ^{15}N-THI5p, This suggested the possibility that THI5p was serving as the donor of the histidine derived atoms of HMP-P.

To eliminate the possibility that HMP-P formation was dependent on the copurification of a tightly bound histidine or histidine metabolite, THI5p was treated with 8 M urea until no activity remained and then renatured by slow dialysis. The resulting protein, purified by Ni-NTA chromatography, showed 59% of the specific

activity of the original preparation (Figure 3.5). Since denaturation destroys the active site structure, and dialysis removes any released metabolite, this experiment provides strong support for the hypothesis that the histidine derived N=C-N fragment is extracted from the THI5 protein.

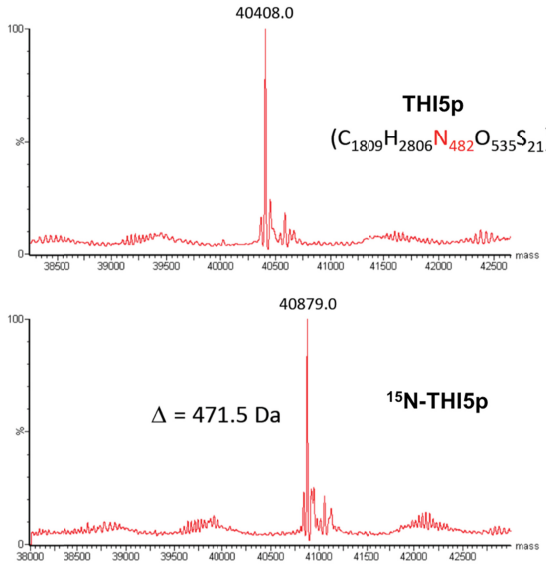


Figure 3.4. The ESI-MS analysis of ^{14}N-THI5p and ^{15}N-THI5p.

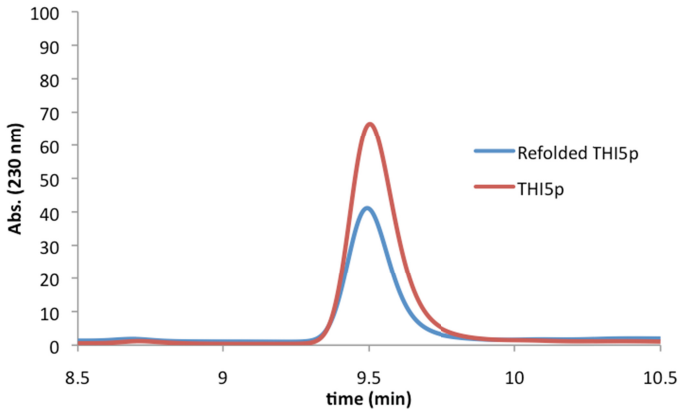
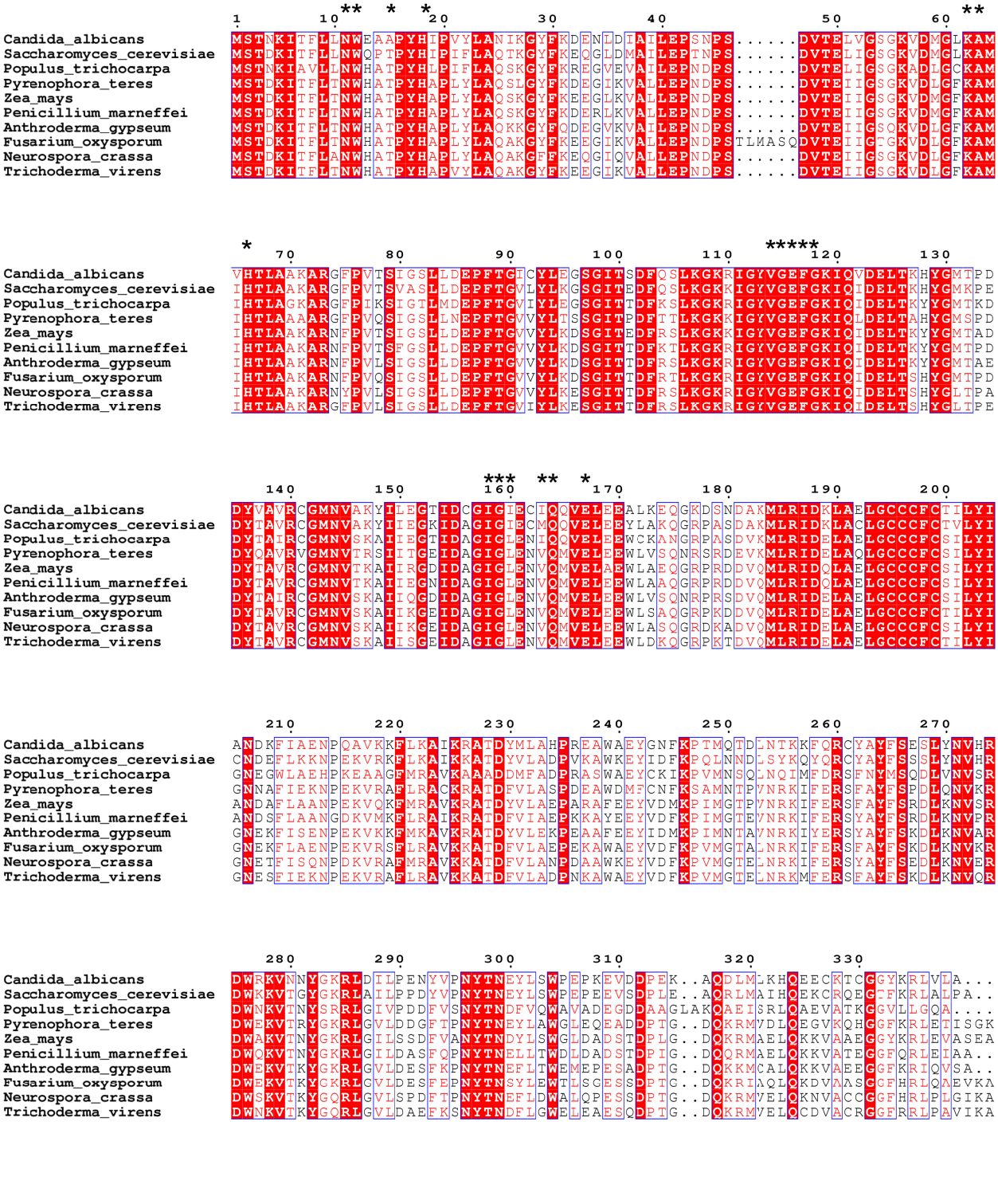


Figure 3.5. The THI5p denatured in urea and refolded is active. Red trace: HMP produced by THI5p before urea denaturation. Blue trace: HMP produced by THI5p after denaturation in 8 M urea and renaturation by slow dialysis.

The aligned THI5p sequences from *C. albicans*, *S. cerevisiae*, *Neurospora crassa* and *Schizosaccharomyces pombe* are shown in Figure 3.6. The *C. albicans*

Figure 3.6. Sequence alignment for THI5p. Ten diverse sequences were selected. Conserved residues are highlighted in red. Residues that line the active site cavity are indicated by asterisks.


Candida_albicans
Saccharomyces_cerevisiae
Populus_trichocarpa
Pyrenophora_teres
Zea_mays
Penicillium_marneffei
Anthroderma_gypseum
Fusarium_oxysporum
Neurospora_crassa
Trichoderma_virens

THI5p sequence has six histidine residues. Of these, His18 and His66 are absolutely conserved in all species. If one of these histidine residues is the N=C-N donor, we would expect THI5p, after HMP-P formation (inactive THI5p) to show a mass deficit. However MS analysis of a tryptic digest of inactive THI5p gave fragments containing only five of the six-histidine residues; we were unable to identify a peptide containing His66 or modified His66 in samples of active and inactive enzyme. We do not yet understand why the MS analysis failed to identify this peptide (Figure 3.7). Each of the six histidine residues was then mutated in an alternative attempt to identify the histidine donor for HMP-P formation. The H128N, H234N, H273N and H323N mutants were active while the H18N mutant was inclusion body, and H66N mutant was inactive (Table 3.3). This suggests that one of the two absolutely conserved histidine residues might be the source of the N=C-N fragment.

Table 3.3. THI5p histidine mutants and their activities

Mutant		**Mutant**	
H18N	Inclusion body	H18D	Inclusion body
H66N	Inactive	H18A	Inclusion body
H128N	Active	H66G	Inactive
H234N	Active		
H273N	Active		
H323N	Active		

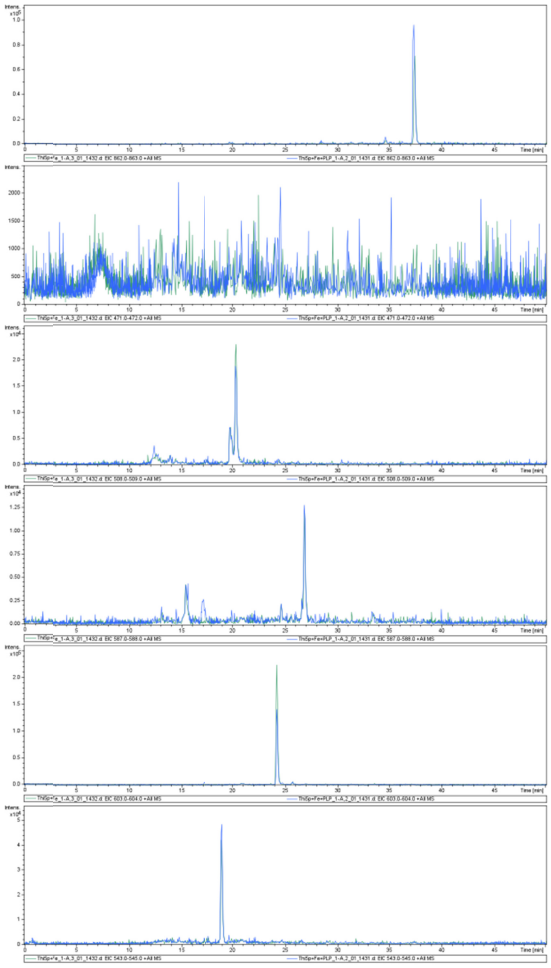


Figure 3.7. LC-MS analysis (extracted ion chromatograms) of each of the six histidine-containing tryptic peptides derived from active and inactive THI5p. Blue trace: peptides derived by trypsinolysis of the THI5p + Fe(III) + PLP reaction mixture. Green trace: peptides derived by trypsinolysis of the THI5p + Fe(III) control. All expected histidine containing peptides were detected except for the peptide spanning residues 63-71.

Crystal Structure of THI5p Complexed with PLP. X-ray crystallography of the *C. albicans* THI5p complexed to PLP provided further insight into the N=C-N source. The active site of this enzyme (Figure 3.8) shows PLP bound as an imine with Lys62 and His66 stacked on top of it. On the opposite face of the PLP is a cavity lined with

conserved residues, presumably providing the iron binding site and the acid base residues needed to catalyze the conversion of a His-PLP adduct to HMP-P.

Early on it was apparent that residues beyond Pro307 were missing for both protomers in the CaTHI5p/PLP model, presumably as the result of a secondary cleavage by thrombin. However, these residues were not involved in crystal packing contacts and were far away from the active site. Reexamination of the THI5p-H66G structure showed that the B-factors for residues beyond Pro307 were about twice those for the rest of the protein, suggesting that the crystal structure contained a mixture of truncated and full length protein.

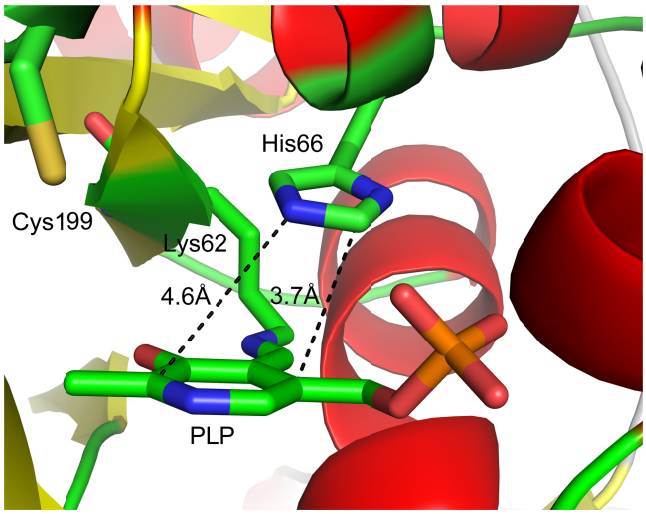


Figure 3.8. Active site of the *C. albicans* THI5p showing PLP bound via an imine to Lys62 and His66 in close proximity to the PLP.

Mechanistic Proposal. A starting mechanistic proposal for the conversion of PLP and histidine to HMP-P is outlined in Figure 3.9. In this mechanism the active site PLP imine **13** undergoes a tautomerization to give **14**. The azadiene part of **14**, activated by electron donating and electron withdrawing substituents, undergoes a formal Diels-Alder reaction with the CN double bond of His66 to give **15**.

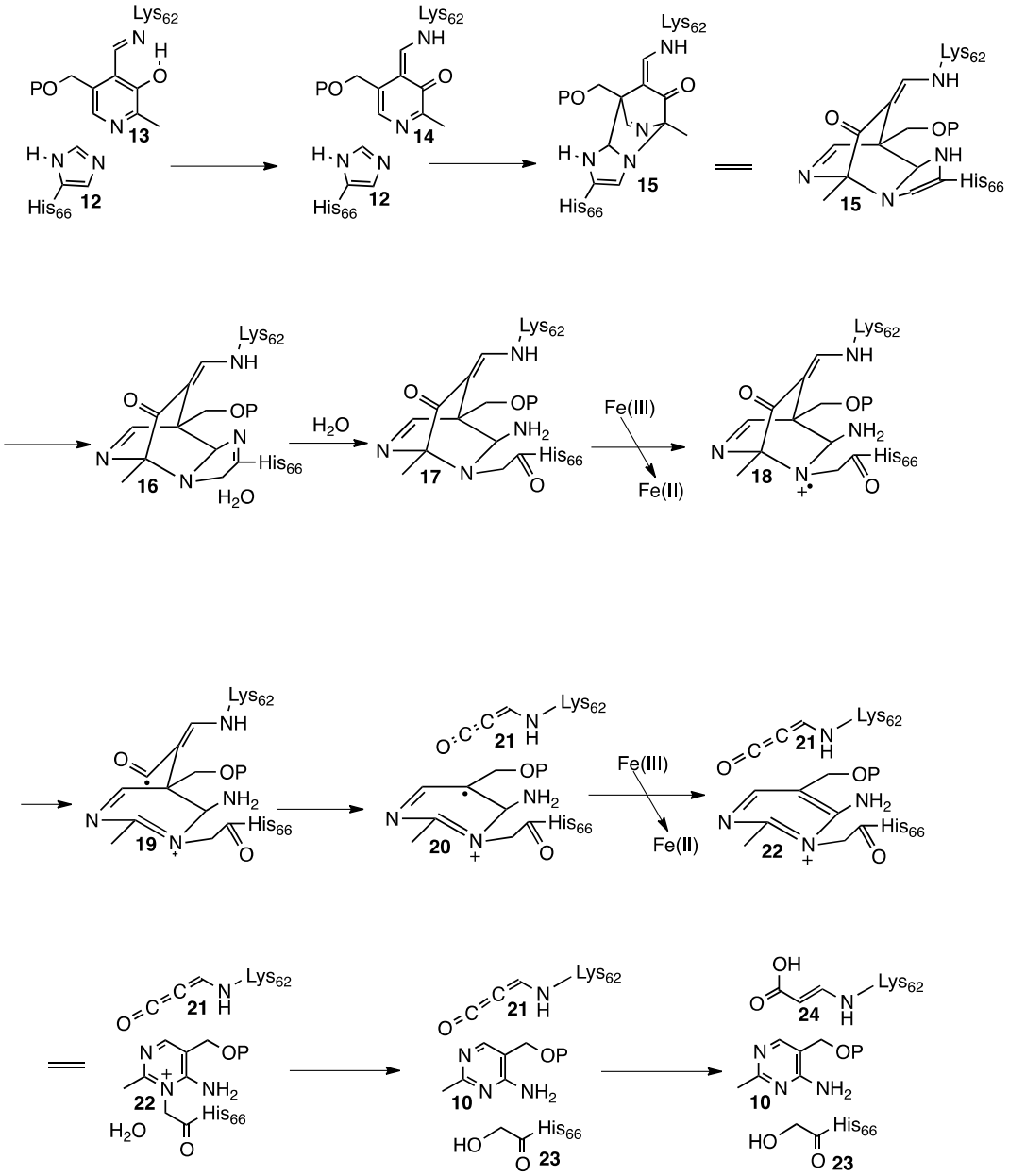


Figure 3.9. Mechanistic proposal for the formation of HMP-P from PLP and the active site His66.

Tautomerization to **16** followed by imine hydrolysis gives **17**. Fe(III) mediated oxidation of the tertiary amine of **17** gives **18,** which can then undergo two radical mediated β-scission reactions to give **20**. After oxygen mediated regeneration of Fe(III), oxidation of **20** gives **22**. A substitution reaction, possibly mediated by Cys199 releases the HMP-P **10** from the protein. Hydrolysis of heterocumulene **21** to

24 completes the reaction. Many features of this proposal are readily testable. For example, Extended X-ray Absorption Fine Structure and Electron Paramagnetic Resonance Spectroscopy can be used to identify the iron binding site coordination and the oxidation state of iron. Different mutants and reactions can be introduced to explore different intermediate conditions throughout the reaction and elucidated using Mass Spectrometry and X-ray Crystallography. Homologs of THI5p from other species can also be introduced to systematically study this suicidal behavior of the protein.

In conclusion, active THI5p can be readily overexpressed in *E. coli*. The protein uses PLP and the active site His66 to form HMP-P. This reaction requires oxygen and Fe(III) and generates inactive THI5p. A structure of *C. albicans* THI5p shows PLP bound at the active site via an imine with Lys62 and His66 in close proximity to the PLP. A mechanistic proposal for a reaction between His66 and PLP to generate HMP-P is described and demonstrated the unprecedented chemistry for this suicidal enzyme. In this reaction, THI5p not only serves as an enzyme, but also a co-substrate for the formation of HMP-P and becomes inactive after a single turnover. This suicidal behavior of enzymes is exceedingly rare in nature (*24, 25*) and the biological function of this modified enzyme, if any, has not yet been revealed.

REFERENCES

1. Jurgenson, C. T., Begley, T. P., and Ealick, S. E. (2009) The structural and biochemical foundations of thiamin biosynthesis, *Annu. Rev. Biochem. 78*, 569-603.
2. Begley, T. P., Chatterjee, A., Hanes, J. W., Hazra, A., and Ealick, S. E. (2008) Cofactor biosynthesis--still yielding fascinating new biological chemistry, *Curr. Opin. Chem. Biol. 12*, 118-125.
3. Chatterjee, A., Hazra, A. B., Abdelwahed, S., Hilmey, D. G., and Begley, T. P. (2010) A "Radical Dance" in Thiamin Biosynthesis: mechanistic Analysis of the Bacterial Hydroxymethylpyrimidine Phosphate Synthase, *Angew. Chem., International Edition 49*, 8653-8656, S8653/8651-S8653/8615.
4. Chatterjee, A., Abeydeera, N. D., Bale, S., Pai, P.-J., Dorrestein Pieter, C., Russell David, H., Ealick Steven, E., and Begley Tadhg, P. (2011) *Saccharomyces cerevisiae* THI4p is a suicide thiamine thiazole synthase, *Nature 478*, 542-546.
5. Zeidler, J., Sayer, B. G., and Spenser, I. D. (2003) Biosynthesis of vitamin B1 in yeast. Derivation of the pyrimidine unit from pyridoxine and histidine. Intermediacy of urocanic acid, *J. Am. Chem. Soc. 125*, 13094-13105.
6. Zeidler, J., Ullah, N., Gupta, R. N., Pauloski, R. M., Sayer, B. G., and Spenser, I. D. (2002) 2'-hydroxypyridoxol, a biosynthetic precursor of vitamins B(6) and B(1) in yeast, *J. Am. Chem. Soc. 124*, 4542-4543.
7. Tazuya, K., Azumi, C., Yamada, K., and Kumaoka, H. (1995) Pyrimidine moiety of thiamin is biosynthesized from pyridoxine and histidine in *Saccharomyces cerevisiae*, *Biochem. Mol. Biol. International 36*, 883-888.
8. Himmeldirk, K., Sayer, B. G., and Spenser, I. D. (1998) Comparative Biogenetic Anatomy of Vitamin B1: A 13C NMR Investigation of the

Biosynthesis of Thiamin in *Escherichia coli* and in *Saccharomyces cerevisiae*, *J. Am. Chem. Soc. 120*, 3581-3589.

9. Ishida, S., Tazuya-Murayama, K., Kijima, Y., and Yamada, K. (2008) The direct precursor of the pyrimidine moiety of thiamin is not urocanic acid but histidine in *Saccharomyces cerevisiae*, *J. Nutr. Sci. Vitaminol. 54*, 7-10.
10. Wightman, R., and Meacock, P. A. (2003) The THI5 gene family of *Saccharomyces cerevisiae*: distribution of homologues among the hemiascomycetes and functional redundancy in the aerobic biosynthesis of thiamin from pyridoxine, *Microbiology 149*, 1447-1460.
11. Seckute, J., McCloskey, D. E., Thomas, H. J., Secrist, J. A., 3rd, Pegg, A. E., and Ealick, S. E. (2011) Binding and inhibition of human spermidine synthase by decarboxylated S-adenosylhomocysteine, *Protein Sci. 20*, 1836-1844.
12. Dauter, Z., Dauter, M., and Rajashankar, K. R. (2000) Novel approach to phasing proteins: derivatization by short cryo- soaking with halides, *Acta Crystallogr. D 56*, 232-237.
13. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, In *Methods Enzymol.* (Carter, C. W. J., and Sweet, R. M., Eds.), pp 307-326, Academic Press, New York.
14. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD, *Acta Crystallogr. D 58*, 1772-1779.
15. Sheldrick, G. M. (2008) A short history of SHELX, *Acta Crystallogr. A 64*, 112-122.
16. Pape, T., and Schneider, T. R. (2004) HKL2MAP: a graphical user interface for phasing with SHELX programs, *J. Appl. Cryst. 37*, 843-844.
17. Collaborative Computational Project-Number 4. (1994) The CCP-4 suite: programs for protein crystallography, *Acta. Crystallogr. D 50*, 760-763.

18. Zhang, K. Y., Cowtan, K., and Main, P. (1997) Combining constraints for electron-density modification, *Methods Enzymol. 277*, 53-64.

19. Terwilliger, T. C. (2003) Automated main-chain model building by template matching and iterative fragment extension, *Acta Crystallogr. D 59*, 38-44.

20. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr. D Biol. Crystallogr. 66*, 486-501.

21. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr. D Biol. Crystallogr. 67*, 355-367.

22. Gopaul, D. N., Meyer, S. L., Degano, M., Sacchettini, J. C., and Schramm, V. L. (1996) Inosine-uridine nucleoside hydrolase from *Crithidia fasciculata*. Genetic characterization, crystallization, and identification of histidine 241 as a catalytic site residue, *Biochemistry 35*, 5963-5970.

23. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283-291.

24. Sedgwick, B., Robins, P., Totty, N., and Lindahl, T. (1988) Functional domains and methyl acceptor sites of the Escherichia coli ada protein, *J. Biol. Chem. 263*, 4430-4433.

25. Chatterjee, A., Abeydeera, N. D., Bale, S., Pai, P. J., Dorrestein, P. C., Russell, D. H., Ealick, S. E., and Begley, T. P. (2011) Saccharomyces cerevisiae THI4p is a suicide thiamine thiazole synthase, *Nature 478*, 542-546.

CHAPTER 4

OXIDATIVE MODIFICATION OF CYSETINE INTO SULFONIC ACID IN SUICIDE ENZYME THIAMIN PYRIMIDINE SYNTHASE

Section 4.1. Introduction

Thiamin pyrophosphate is an essential cofactor in all living systems. While bacteria, archea, and lower eukaryotes produce their own thiamin, higher eukaryotes cannot and require thiamin as a dietary supplement. Deficiency of thiamin in human can cause beriberi (*1*) and WenickeKorsakoff syndrome (*2*). In both prokaryotes and eukaryotes the thiazole and pyrimidine moieties are biosynthesized separately, and then coupled by thiamin phosphate synthase. While prokaryotic biosynthesis of the thiazole moiety requires five gene products, eukaryotes require only one enzyme, THI4p, which we have shown to be a suicide enzyme (*3*). THI4p cysteine residue serves as the thiazole sulfur source, and after one turnover the enzyme is inactive. In prokaryotes and plants, biosynthesis of the pyrimidine moiety utilizes ThiC, a radical S-adenosylmethioninine enzyme that catalyzes the complex rearrangement of 5-aminoimidazole ribonucleotide to form 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMP-P) (*4*). In eukaryotes, HMP-P is biosynthesized from histidine and pyridoxal 5′-phosphate (PLP) and utilizes the enzyme THI5p. Our previous biochemical and structural studies in THI5p from *Candida albicans* have demonstrated that THI5p is also a suicide enzyme (*5*). A histidine residue of THI5p is the source of one carbon and two nitrogen atoms to complete the biosynthesis of HMP-P, and after one turnover THI5p is inactive. A mechanism involving Diels-Alder reaction between the histidine and pyridoxal phosphate (PLP) has been proposed (*5*). Very recently another group crystallized THI5p in *Saccharomyces cerevisiae*, also showing the same PLP binding site as that in *C. albicans*.

Sequence analysis show that the THI5p family has a highly conserved CCCXC motif at the C-terminus. The structures of this motif have also been shown in several crystal structures (*5, 6*). In the structure of THI5p in *S. cerevisiae*, residues 187-195 are prone to disorder as indicated by missing electron density. Coupled with complementary assay for individual cysteine residue, it has been proposed that the cysteine-rich motif coordinates an iron atom for the HMP-P synthesis reaction (*6*). The diverse family of cysteine-rich motifs are often related to the iron and cobalt coordination and frequently serves as the Lewis acid in the electron transfer and radical reactions. For example, the CX_3CX_2C motif in radical-SAM enzymes in *E. coli* serves as the binding site for [4Fe-4S] cluster (*7*). A model of THI5p based on THIYp, a sequence homolog with 25% sequence identity, has also been proposed with a putative [Fe-S] cluster binding site (*8*). However, no experimental evidence of [Fe-S] cluster has been found in the structure of THI5p. Instead, evidence shows that the activity of THI5p depends on Fe(III) ion (*5*), which leads to the hypothesis that the enzyme catalyzes an Fe(III)-mediated redox reaction. The binding of free iron by cysteine has been discovered in other proteins such as nitrile hydratases (*9*). Iron-containing nitrile hydratase share a conserved CX_2CXC motif. This motif serves as part of the octahedral binding site of iron, which consists of the side chains of three cysteines and main chain amides of another two residues (*10*). These ligands provide five vertices of an octahedron for iron, with the sixth position unoccupied and open to the active site cleft.

So far, several structures of THI5p complexed with the substrate PLP have been determined in fungus (*5*) and yeast (*6*). Here we present the comparisons for three new structures of THI5p to the published structure THI5p/PLP complex in *C. albicans*. The three new structures of THI5p include the modified-THI5p, the C199A mutant complexed with PLP and the C-terminus truncated THI5p, E309Ter,

complexed with PLP. We also provide a comprehensive and detailed structural analysis for the previously reported THI5p/PLP complex. Along with the support from the LC-MS data, the comparisons for the four structures provide further insights into the mechanism of this single-turnover enzyme that is without chemical or biochemical precedent.

Section 4.2. Materials and Methods

Cloning and Mutagenesis. The *THI5* gene from *C. albicans* (CaTHI5) genomic DNA was PCR amplified and the purified PCR product was digested and ligated into a pET28 vector as described previously (*5*). The vector was then sub-cloned into pTHT, a pET-28 derived vector for mutagenesis. The site-directed mutagenesis was performed for generating a C199A mutant and a C-terminus truncated mutant E309Ter using a standard PCR-based protocol (*11*) and corresponding forward and reverse primers.

Expression, Purification and Crystallization of Modified-THI5p, C199A/PLP complex and E309Ter/PLP complex. The basic expression and purification procedures were the same as described previously (*5*). After the basic steps of purification, the wild type THI5p and the C199A and E309Ter mutants were treated with iron and PLP following the same procedure as for reconstitution of THI5p activity. 1.5 equivalent of $Fe(NH_4)_2(SO_4)_2$ were added in an ice bath for 30 min. Then the mixture was desalted using an Econo-Pac 10DG desalting column (Bio-RAD) pre-equilibrated with 20 mM NaCl, 10 mM Tris, pH 7.5, 3 mM DTT. Subsequently, 2 equivalent of PLP was added at 22 °C. The reaction mixture was then incubated at room temperature for 3 h and concentrated to 15 mg/mL using an Amicon concentrator (30 kDa MWCO filter, Millipore), flash frozen and stored at -80 ºC.

The initial crystallization condition was determined using sparse matrix screen Crystal Screen 1 and 2 (Hampton Research), Wizard III and IV (Emerald Bio). The

optimized reservoir conditions were 2.5 M NaCl, 0.1 M Na/K phosphate at pH 6.2 for modified-THI5p, 20% (w/v) PEG3350 and 0.2 M NH_4Cl for C199A/PLP complex, 18% PEG 3350 and 0.2 M NH_4Cl for E309Ter/PLP complex. The drops contained 1 μL of protein solution and 1 μL of reservoir solution. Rod-shaped crystals for C199A/PLP and E309Ter/PLP grew within 2 days to the size of 800 μm × 20 μm × 20 μm. Tetragonal bipyramidal shape of modified-THI5p grew after five months to the size of 80 μm × 50 μm × 50 μm. The crystals of the modified-THI5p were briefly soaked in 5% ethylene glycol prior to vitrification in liquid nitrogen, and in 15% glycerol for C199A/PLP and E309Ter/PLP complex.

Data Collection and Structure Determination. X-ray diffraction data of modified-THI5p and mutants of THI5p were collected at the NE-CAT beamline 24-ID-E at APS using an CCD-based ADSC Quantum 315 detector (Area Detector Systems Corporation) at a wavelength of 0.9792 Å with 1 s exposure times and 1° oscillation angles. The detector distances were 360 mm, 250 mm, 330 mm for the modified-THI5p, C199A/PLP, E309Ter/PLP complex, respectively. Data were indexed, integrated and scaled using the HKL2000 program suite (*12*).

The structure of C199A/PLP complex and E309A/PLP complex were determined using the previously solved structure of THI5p/PLP complex as the initial model and refining with PHENIX (*13*). The structures of the modified-THI5p and E309A/PLP complex, which crystallized in different space groups, were determined by molecular replacement using MOLREP (*13*) with a monomer from the refined THI5p structure as the search model. Data collection and refinement statistics are summarized in Table 4.1 and Table 4.2.

Table 4.1. Data Collection Statistics

	THI5p/PLP	Modified-THI5p	C199A/PLP	E309Ter/PLP
beamline	APS 24-ID-C	APS 24-ID-E	APS 24-ID-E	APS 24-ID-E
wavelength (Å)	0.9795	0.9792	0.9792	0.9792
space group	$P2_12_12_1$	$P4_32_12$	$P2_12_12_1$	$P22_12_1$
a (Å)	55.0	89.6	54.9	55.4
b (Å)	99.7	89.6	100.6	115.0
c (Å)	123.4	96.2	126.5	120.4
protomers per asymmetric unit	2	1	2	2
resolution range (Å)	50.0-2.20 (2.34-2.20) [a]	50.0-1.97 (2.04-1.97)	50.0-1.70 (1.76-1.70)	50-2.34 (2.42-2.34)
total no. of reflections	182,558	247,765	272,455	149,464
no. of unique reflections	36,139	28,391	77,607	33,197
redundancy	5.1 (5.0) [a]	8.8 (5.4)	3.7 (3.4)	4.7 (4.2)
completeness (%)	98.9 (99.8) [a]	99.1 (93.4)	94.6 (75.5)	96.8 (94.5)
R_{merge} (%) [b]	10.5 (69.0) [a]	7.9 (48.5)	4.2 (38.1)	8.9 (41.9)
I/σ(I)	14.5 (2.3) [a]	28.2 (2.3)	15.6 (2.1)	10.9 (2.1)

[a]Values in parentheses are for the highest resolution shell.
[b]$R_{merge} = \Sigma\Sigma_i | I_i - \langle I \rangle | / \Sigma \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices h, k, l.

Table 4.2. Data Refinement Statistics

	THI5p/PLP	Modified-THI5p	C199A/PLP	E309Ter/PLP
no. of reflections	36,004	28,082	73,352	32,086
no. of reflections in working set	34,157	26,723	69,667	30,498
no. of protein atoms	4,929	2,303	5,232	4,806
no. of ligand atoms	46	0	46	46
no. of water atoms	318	134	864	420
rmsd for bonds (Å)	0.007	0.008	0.006	0.008
rmsd for angles (°)	1.061	1.375	1.079	1.091
R_{work}/ R_{free}[a] (%)	17.4/22.8	18.5/20.9	15.4/18.4	16.6/21.2
Ramanchandran analysis[b]				
most favored (%)	92.7	95.0	93.1	92.9
additional allowed (%)	7.3	5.0	6.9	7.1
generously allowed (%)	0.0	0.0	0.0	0.0
disallowed (%)	0.0	0.0	0.0	0.0

[a] $R_{work} = \sum_{i=1}^{N_U} \left| |F_{O,i}| - k|F_{C,i}| \right| / \sum_{i=1}^{N_U} |F_{O,i}|$, where $|F_{O,i}|$ and $|F_{C,i}|$ are the observed and calculated structure factor amplitudes for reflection $i = hkl$ and N_U is the number of unique reflections. For R_{free}, these sums were taken over a 5% subset of the reflections excluded during structure refinement. [b] Carried out using PROCHECK (excludes glycine, proline, and end residues) (*14*).

LC-MS Analysis on THI5p and the Mutants. The procedures of trypsin digestion and LC-MS analysis are similar as described previously (5). In LC-MS, following linear gradient, at a flow rate of 0.4 mL/min, on a Synergi 2.5u polar-RP 100A (50X2 mm) was used: solvent A is 0.1% formic acid in water, solvent B is 0.1% formic acid in acetonitrile; 0 min: 100% A; 1 min: 100% A; 8min: 88% A, 12% B; 31min: 65% A, 35%B; 46min: 35% A, 65% B; 48min: 100% B; 53min: 100%B; 56min: 100%A; 65min: 100% A. The HPLC is coupled with a microTOF-Q II for MS analysis in the positive mode.

Section 4.3. Results

Overall Structure of THI5p. THI5p is a homodimer with one dimer per asymmetric unit in the THI5p/PLP complex (Figure 4.1 A). Of the 339 residues, the electron density for the 32 residues at the C-terminus of both protomers was missing in all the native THI5p data sets collected from *C. abicans*. THI5p consists of two domains each containing a five-stranded mixed β-sheet (Figure 1B). The central β-sheet of domain I shows the topology of 2↑1↑3↑10↓4↑, flanked by three α-helices and two 3_{10}-helices on one side, and seven α-helices and one 3_{10}-turn on the other. Domain II, albeit smaller, shows a similar pattern of the β-sheet 7↑6↑8↑5↓9↑, with four α-helices and two 3_{10}-helices on one side and two α-helices and one 3_{10}-helix on the other (Figure 1C). Domain I and II are related by pseudo-2-fold symmetry and joint by two crossovers: one joining β4 and β5 and the other joining η6 and α9. Such central β-sheet arrangement is characteristic for the Type II periplasmic binding protein (PBP), featured by the central β-sheet topology of 2↑1↑3↑n↓4↑ in both domains, where n stands for the strand right after the first cross-over from the N-terminus towards the C-terminus, and *vice versa (15)*. The overall structure of the protomer superimposes fairly well with the predicted homology model (*8*), generated using periplasmic N-formyl-4-amino-5-(aminomethyl)-2-methylpyrimidine (FAMP)

binding protein ThiY as the template, which shares 25% sequence identity and 47% similarity to THI5p. In spite of variance in the number of protomers per asymmetric unit, THI5p from *C. albicans* is always shown to be a dimer built from the same interface. The interface is formed by 16 hydrogen bonds and 5 salt bridges highly conserved in all structures, according to PISA (*16-18*). Similar interface is also observed in the structures of THI5p from *S. cerevisiae* (*6*).

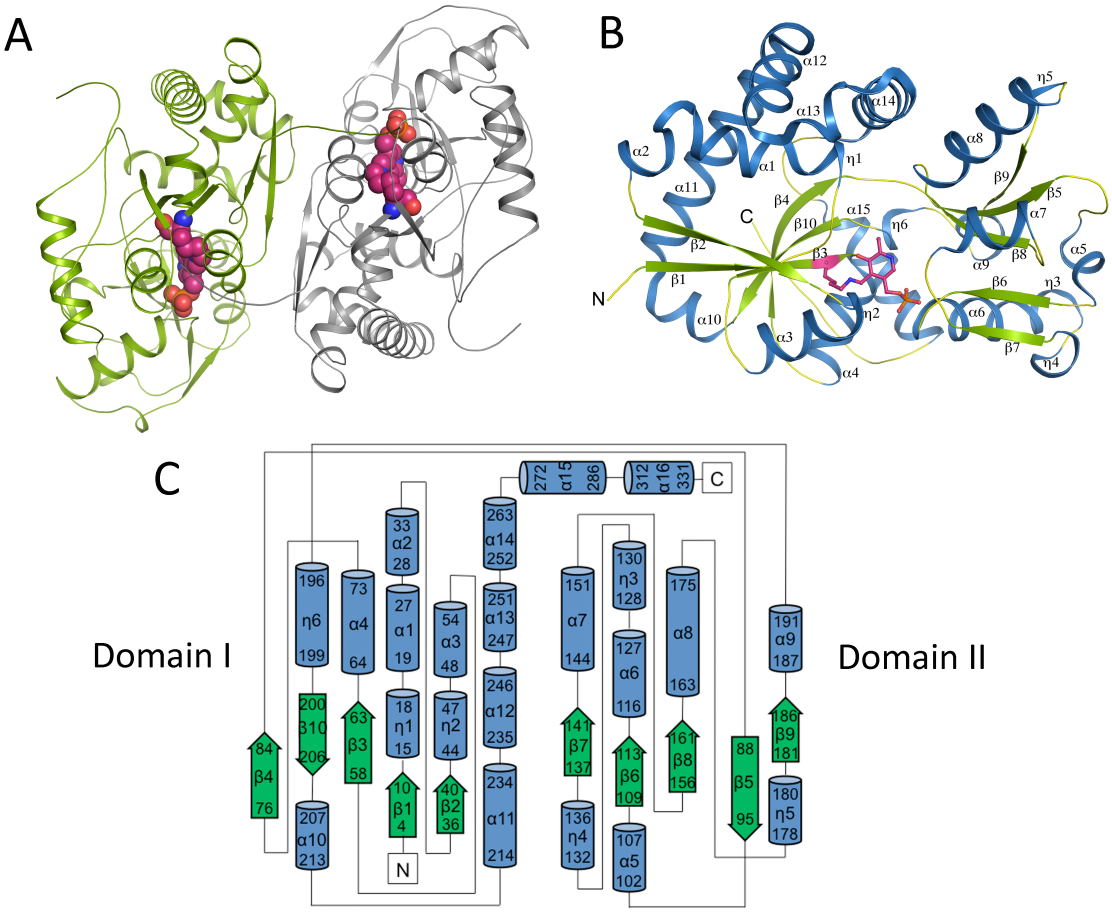


Figure 4.1. Overall structure of THI5p. (A) Ribbon diagram of THI5p. The protomers are colored green and grey. The PLP in the active site is indicated by the space filling model in pink. (B) Ribbon diagram of THI5p protomer is color coded by α-helices and 3_{10}-helices (blue) as well as β-strands (green) are labeled. (C) Topology diagram of THI5p. The blue cylinders represent α-helices and 3_{10}-helices and the green arrows represent β-strands. The first and last residue number of each secondary structural element is indicated.

THI5p Active Site. The active site, as defined by the structure of the THI5p/PLP complex, is positioned in the cleft between the two domains (Figure 4.1 B). Domain I serves as the pyridoxal binding domain and Domain II as the phosphate binding domain. PLP is anchored in the active site via the imine formed between the 4′-aldehyde group and the N6-amine group of Lys62, with His66 stacked on top of it (*5*). The distance between the pyridine ring and the histidine imidazole ring is within the van der Waals distance (~4 Å), thus unveiling the identity of the catalytic histidine residue (Figure 4.2 A). The 3-hydroxyl group of PLP forms hydrogen bond with Asn11 located in the loop region between β1 and η1. Below the pyridine ring, Trp12 is positioned in parallel to form π-π interactions, thereby further stabilizing the pyridine ring. In the phosphate binding site, phosphate binding takes place through H-bonds to the backbone amides of Gly115, Glu116, Phe117 and Gly118. These residues form the N-terminus of a long α-helix and the helix-dipole is positioned to stabilize phosphate binding. At the other side of the active site, Cys199 is about 5 Å away from the pyridoxal ring (Figure 4.2 B), bridging the active site to a cysteine-rich motif, CCCFC (residue 194-199). The cysteines are located in η6 that are tightly constrained between α9 and β10. While the sulfhydryl of Cys199 points towards PLP, sulfhydryls of Cys195 and Cys197 pointed towards each other, with a distance of 3.0 Å. This distance is a bit longer than an average disulfide bond distance of 2.3 Å, indicating a mixture of disulfide and thiol. Cys196 points away from Cys195 and Cys197, towards a salt bridge formed by Asp188 and Arg186. Despite the fact that cysteine-rich motif is often involved in [Fe-S] cluster binding, no evidence of [Fe-S] cluster has been found in THI5p (results not shown). Nevertheless, the protein activity towards the binding of mono iron has been confirmed (*5*), although no convincing electron density for iron has be identified in the structure. Considering the close

proximity of Cys199 to PLP, the CCCFC motif could be potentially involved in iron-mediated electron transfer and radical reactions (*7*).

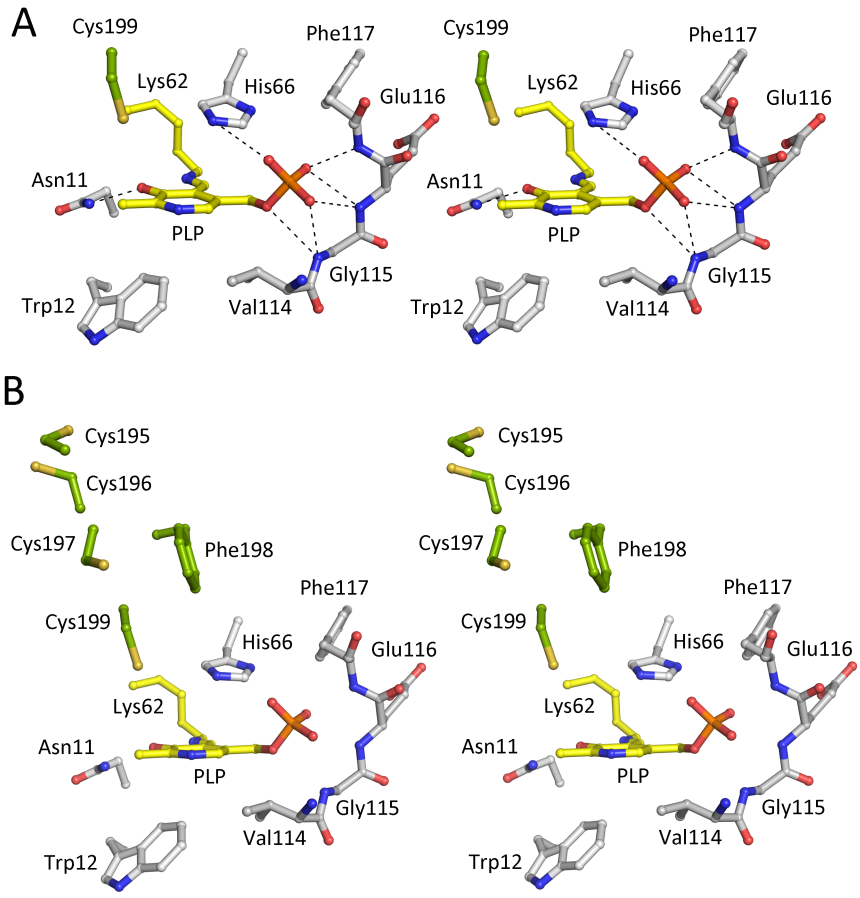


Figure 4.2. Active site of THI5p/PLP complex. (A) Stereoview of the active site residues of THI5p complexed with covalently bound PLP. Active site residues within van der Waals radii of the PLP molecules are shown. Hydrogen bonds formed between PLP and amino acid side chains are represented by black dashes. (B) Stereoview of the CCCFC motif in proximity to the PLP binding site. Protein atoms are colored grey and ligand atoms are colored yellow.

The Structure of Modified-THI5p. After incubating THI5p with $Fe(NH_4)_2(SO_4)_2$ and PLP, tetragonal bipyramidal shaped crystals appeared after five months. Instead of the orthorhombic space group $P2_12_12_1$ as the THI5p/PLP complex, the modified-THI5p was crystallized in tetragonal space group $P4_32_12$, in a higher

symmetry. The crystal packing became tighter and involved more protein-protein contacts while retaining the same dimer interface.

In each protomer, Domain I maintains the same geometry as the pre-reaction complex, whereas Domain II undergoes significant conformational changes (Figure 4.3) according to Difference Distance Matrix Analysis (*20*). Contrary to the hinge-

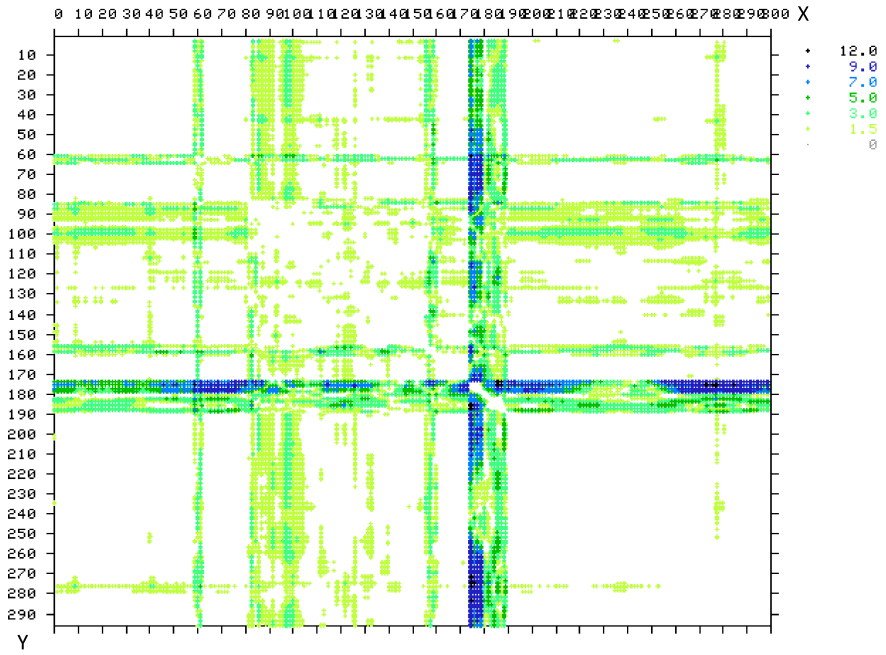


Figure 4.3. Difference Distance Matrix Analysis of THI5p/PLP complex (X-axis) vs. modified-THI5p (Y-axis).

bending motion that has been observed in Domain II (equivalent to the phosphate binding domain) of Type II PBPs that undergo a 50° rotation and a negligible transition (*21*) or twist (*22*), the secondary structure in THI5p is mostly engaged in local movements (Figure 4.5 A). Starting from the two crossovers, η6 and the loop between β4 and β5 change the conformation substantially. The backbone of helix-η6 stretches out, connects to the backbone of residues 86-88 through hydrogen bonds and forms an anti-parallel β-sheet. Notably, η6 corresponds to residue 196-199, wherein the CCCFC motif is. The three-dimensional arrangement of the CCCFC motif alters dramatically. Instead of having Cys195 and Cys197 pointing into each other, a

disulfide bridge forms between Cys195 and Cys196. The flanking helix containing residues 187-191 also unwinds and becomes part of the loop connecting α9 and η6. The loop rotates for about 40° and the movement propagates to α9 and β9, involving the displacement of the central pleated β-sheet for about 2 Å.

On the other side of Domain II, a small loop between β6 and α6 shifts in for about 2.5 Å. This region corresponds to the phosphate binding site, where residues 115-118 are located and interact with PLP through backbone hydrogen bonding to the phosphate group. In addition, the electron density between surface residues 176 and 182 is missing, corresponding to the 3_{10}-helix η5 in the THI5p/PLP complex.

Although no ligand has been found in the active site, it does not seem to be simply an apo form of the protein due to several difference densities shown in the structure. The density for the Cys199 side chain is significantly larger than a normal thiol group (Figure 4.4 A). When the cysteine was modeled in, difference density showed up at three corners. Instead of the globular-shaped density of a normal sulfhydryl group, the density looks more planar, resembling a triangle from the top down view (Figure 4.4 B), with one dimension slightly longer than the other. Combined with the LC-MS data (Figure 4.6), the map indicates the formation of sulfonic acid.

Besides the altered cysteine density, another two regions remain curious. But due to the poor map quality of some residues, it can be hard to interpret. One is the negative density on His66, which was initially modeled in as histidine. The negative difference density covers half of the imidazole ring at the contour level of 3.0 σ. In the composite omit map, the histidine contains the density for only half the ring, missing the N==C—N corner of imidazole (Figure 4.4 C, D). The other piece of puzzle involves an unknown density on the surface of the protein, connecting the side chains of Cys162 and Cys197 (Figure 4.4E). The spatial distance between the two

cysteine residues is 13 Å. The unknown density links the cysteine side chains through the sulfhydryl group.

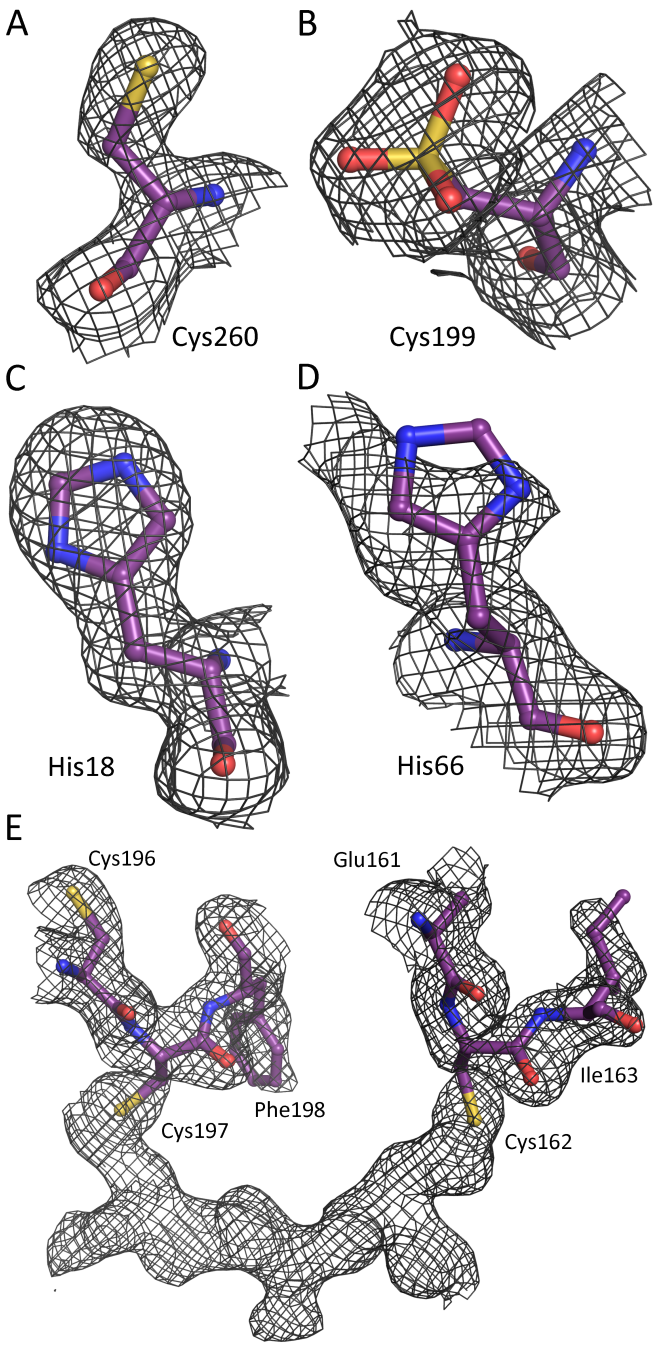


Figure 4.4. Composite omit map for residues in modified-THI5p. The map is contoured at 1σ. (A) Electron density corresponding to a normal cysteine residue, Cys260, in the modified-THI5p. (B) Electron density corresponding to the modified

cysteine residue, Cys199, with the thiol group replaced by sulfonic acid. (C) Electron density corresponding to a normal histidine residue, His18. (D) Electron density corresponding to His66. (E) Electron density between the side chains of Cys162 and Cys197.

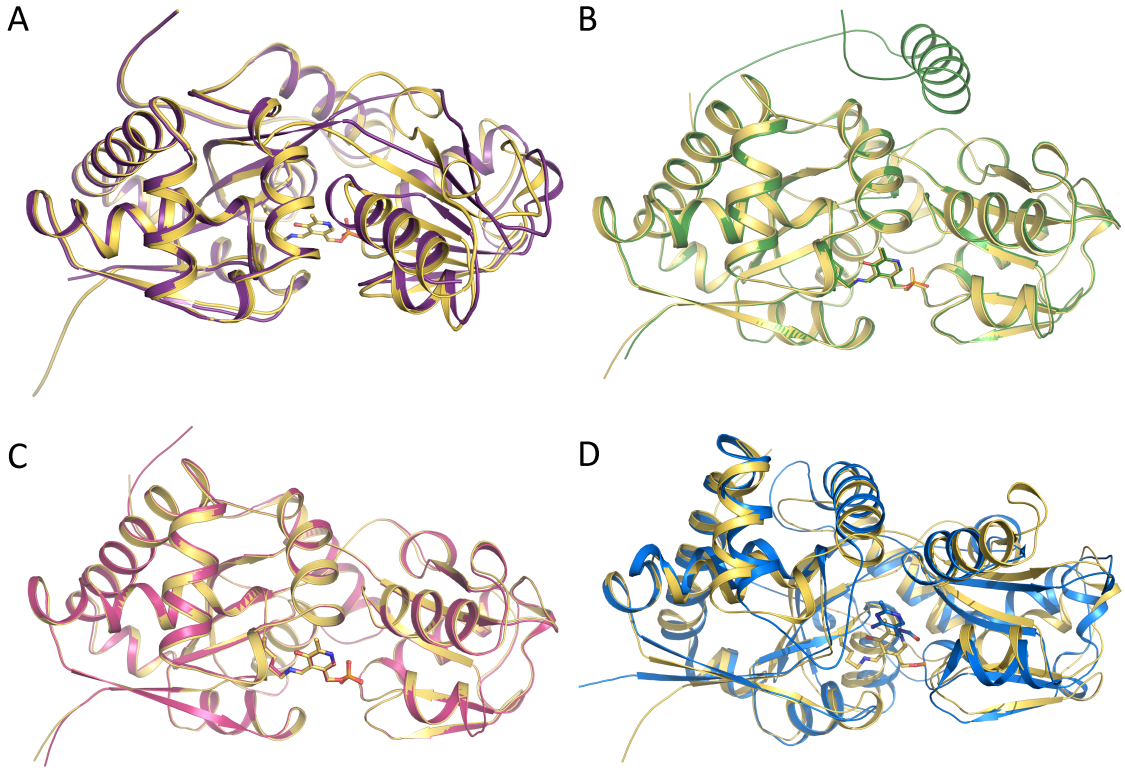


Figure 4.5. The structural comparisons of THI5p/PLP complex (yellow) with (A) the modified-THI5p (purple), (B) the C199A/PLP complex (*19*), (C) the E309Ter/PLP complex (pink), and (D) the THIYp/FAMP complex (blue).

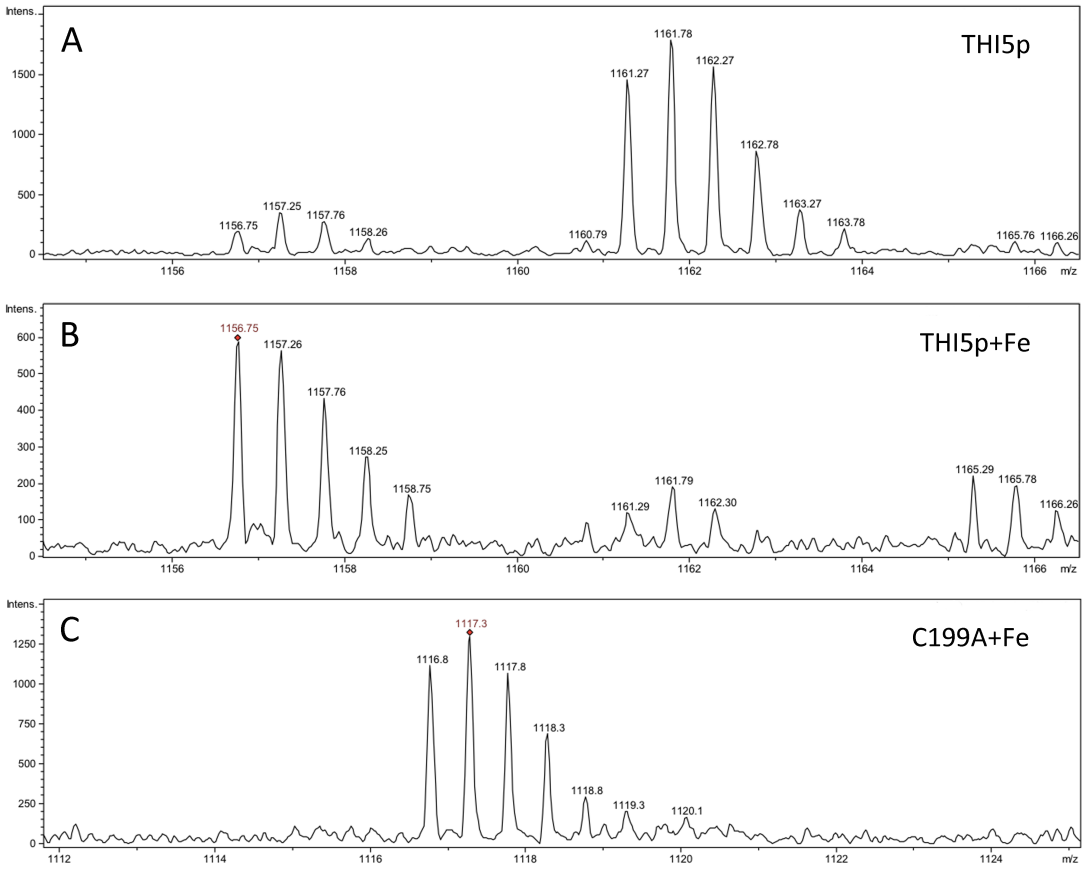
The Structure of C199A/PLP. Following the same procedure as for the modified-THI5p protein, the C199A mutant was cocrystallized with $Fe(NH_4)_2(SO_4)_2$ and PLP and yellow crystals were obtained. One molecule of PLP was observed in the active site for each protomer, retaining the same active site geometry as the native protein (Figure 4.5 B). The reaction did not proceed in spite of the addition of $Fe(NH_4)_2(SO_4)_2$ and PLP. The overall structure of C199A/PLP complex is very similar to the THI5p/PLP complex, with the RMSD of only 0.2 Å, in contrast to 2 Å between the THI5p/PLP complex and the modified-THI5p. The only major difference occurs at the C-terminus. In contrast to the missing density for the C-terminus of the native protein, C199A shows the full length protein with an α-helix at the C-terminus.

Contrary to the overall B-factor of 16.5 $Å^2$ for all atoms, the C-terminal 32 residues have the average B-factor of 29.9 $Å^2$, which is almost twice as much as the rest of the protein, reflecting the flexibility of the C-terminus. Besides the difference at the C-terminus, the distance between the two Sγ of Cys195 and Cys197 (4.0 Å) is much longer than that of the THI5p/PLP complex (2.8 Å).

E309Ter mutant. In order to assess if the presence or absence of the flexible C-terminus can affect the overall structure, E309Ter, the C-terminus truncated version of THI5p was cocrystallized with $Fe(NH_4)_2(SO_4)_2$ and PLP, according to the same procedure as the C199A/PLP complex. The protein was crystallized in the space group of $P22_12_1$, different from the THI5p/PLP complex. Given the altered crystal packing, the active site also has a PLP molecule bound and the overall dimeric structure is very close to THI5p/PLP complex, with the RMSD of only 0.3 Å (Figure 4.5 C). This further implies that the C-terminus does not affect the overall conformation of THI5p.

Detection of Modified Cys199 by Mass Spectrometry. The site of Cys199 modification was also identified by LC-MS analysis of THI5p following trypsin digestion (Figure 4.6). Trypsin cleaves after lysine and arginine residues. Without adding $Fe(NH_4)_2(SO_4)_2$, trypsin digestion yields a peptide detected by LC-MS, corresponding to the sequence of LAELGCCCFCTILYIANDK (Mass=2320.54 Da, amino acids 190-208), with four cysteine side chains carbamidomethylated by iodoacetamide. With the addition of $Fe(NH_4)_2(SO_4)_2$, the molecular weight of the peptide changed to 2311.50 Da, consistent with the modification of three carbamidomethylation and one trioxidation. The absence of one carbamidomethylation and the addition of trioxidation indicated that one of the cysteine residues was unable to be carbamidomethylated, whereas one residue was trioxidized. These two concurrent events indicated that the missing

carbamidomethylation and the oxidation might have occurred to the same cysteine residue. In contrast, when $Fe(NH_4)_2(SO_4)_2$ were added to the C199A mutant, the LC-MS analysis showed that C199A had only three carbamidomethylations



(Mass=2232.05 Da). This thereby implied that Cys199 was the site of trioxidation modification since the C199A mutant was unable to undergo trioxidation.

Figure 4.6. Analysis of Cys199 by LC-MS. (A) LC-MS data of a peptide from trypsin digestion of apo THI5p containing the CCCFC motif. The observed 2320.54 Da corresponds to the sequence of LAELGCCCFCTILYIANDK (amino acids 190-208). (B) The observed mass of the peptide shifts to 2311.50 with the addition of iron in THI5p. (C) The observed mass of the peptide shifts to 2232.05 Da with the addition of iron in C199A.

Section 4.4. Discussions

THI5p and Periplasmic Binding Proteins. PBPs, together with the ATP-binding cassette transport systems, are involved in the active transport of small

molecules and ions (*23*). Despite the large sequence variance, PBPs share common features in the three-dimensional structure. PBPs generally consist of two globular domains with α/β fold, bisected by a cleft wherein the ligand is buried. The binding of the ligand frequently involves a hinge-binding motion between the two domains (*24*).

According to the topological arrangement, PBPs can be classified into three types by the number of inter-domain connections: Type I has three crossovers, Type II has two and Type III has one (*15, 25*). While little is known about the topology of the Type III PBPs, structures of Type I and Type II have been extensively studied. Both domains in the Type I PBPs have the topology of $\beta_2\beta_1\beta_3\beta_4\beta_5$. In contrary, Type II PBPs has the topology of $\beta_2\beta_1\beta_3\beta_n\beta_4$ in both domains, in which β_n stands for the strand immediately after the crossover from the other domain.

In the case of THI5p, Domain I has the topology of $\beta_2\beta_1\beta_3\beta_{10}\beta_4$ and Domain II of $\beta_7\beta_6\beta_8\beta_5\beta_9$, wherein β_{10} and β_5 are the strands right after the crossover in their respective domain. Therefore, THI5p belongs to Type II PBPs. Since the lysine-PLP adduct is positioned in the groove between the two domains, the ligand binding pattern of THI5p is also consistent with the characteristics of Type II PBPs.

However, in spite of the significant conformational change in the course of ligand binding and release in THI5p, its conformational change differs from the hinge-bending motion of other PBPs. Upon the ligand binding, Domain I retains the same position, whereas Domain II rotates for about 50° (*21*), closing around the bound ligand through a Venus flytrap mechanism (*26*). Although THI5p also has Domain I anchored, and Domain II going through a series of conformational changes, the movements in Domain II are primarily alterations of secondary structures and local rotations and translations.

Comparison with other Structures. DALI search (*27*) results showed that the FAMP-binding protein THIYp (PDB code 3IX1) (*8*) has the highest structural

similarity to THI5p, with the Z score of 33.6 and RMSD of 2.3 Å. Other structural homologs include possible thiamin biosynthesis enzyme (PDB code 2X7P, 2X7Q) (*28*), alkane sulfonate-binding protein SsuA (PDB code 3KSJ, 3KSX, 3E4R and 2X26) (*29*), periplasmic nitrate-binding protein NrtA (PDB code 2G29) (*30*) and biocarbonate transport protein CmpA (PDB code 2I48, 2I49, 2I46 and 2I4C) (*31*). All of these proteins are dimeric periplasmic binding proteins and serve different roles in small molecule transport through membranes.

Although THIYp and THI5p both contain a dimer per asymmetric unit, the packing interfaces are quite different. The THIYp dimer interface consists of β6, β7, α10 and the loops between α6 and β8 (*8*), while the interface in THI5p is composed of β1, β2, η2, α7, α10, α13 and the loop connecting α4 and β4. When superimposing one protomer of THI5p on THIYp, the other protomer of THI5p will be at almost a 180° rotation from that of THIYp. Despite distinct protomer arrangements, the tertiary structures of THI5p and THIYp are very close, with RMSD of 2.3 Å. In the structure of THIYp complexed with FAMP, a thiamin salvage pathway intermediate and a HMP homolog, FAMP is buried in the active site and makes extensive stacking and hydrogen bonding interactions with the environment (*8*). The superposition of THIYp and THI5p structures shows that the backbone traces of the two structures align well and the ligand binding site are in close proximity (Figure 5D). The distance between the FAMP pyrimidine ring in THIYp and the PLP pyridoxal ring in THI5p is about 5.5 Å. FAMP is stacked on top of Trp39 in THIYp, which corresponds to Trp12 in THI5p. Due to the structural similarity between FAMP and HMP-P, the product of THI5p, the binding position of FAMP could therefore imply a putative binding site for HMP-P in THI5p (Figure 4.7).

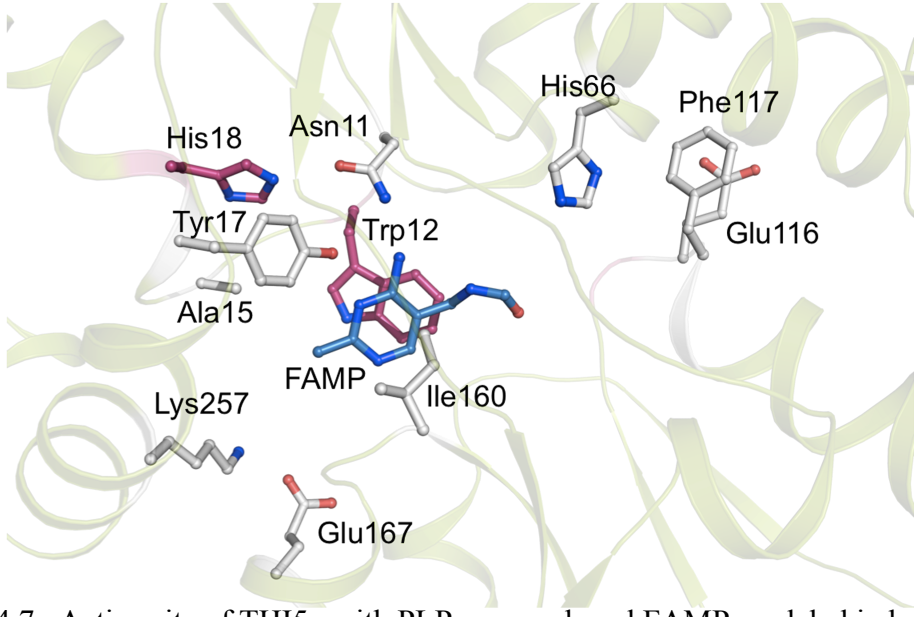


Figure 4.7. Active site of THI5p with PLP removed, and FAMP modeled in based on the superposition of the structures of THI5p/PLP and THIYp/FAMP. The ribbon diagram of THI5p is colored green. The conserved residues shared by both THI5p and THIYp are colored purple. The rest of residues in the THI5p active site are colored grey. The ligand FAMP is colored blue.

The Modified Cysteine Side chain. In the structure of modified-THI5p, apart from the conformational changes, an unusual density was observed in the side chain of Cys199. Cys199 was the last cysteine residue in the CCCFC motif and was absolutely conserved among all THI5p homologs. The side chain of Cys199 showed a much larger electron density than a normal thiol group. This observation was further explained by LC-MS analysis on the peptide containing the CCCFC motif, in which Cys199 was identified as the site for trioxidation into sulfonic acid and thereby was unable to be carbamidomethylated. The electron density for the modified thiol is also consistent with a sulfonic acid. The triangular shaped head of the modified thiol group has one dimension slightly longer than the other, representing the hydroxyl group.

The oxidation on cysteine has been identified in various oxidation states in proteins. Due to the interdependency of the catalytic activity and the wide variety of redox chemistry, many cysteine proteins are considered as "redox switches" to sense the concentration of oxidative stressors (*32*). Cysteine-rich proteins also participate in

the signaling pathways involving proteins such as apoptosis signal-regulated kinase 1 (*33, 34*), bacterial transcription factor OxyR (*35*) and protein tyrosine phosphatases (*36*). Among all the oxidation and reduction reactions, the most common oxidation states for cysteine are the thiol and disulfide. The conversion from thiol to disulfide is reversible in various proteins such as thioredoxin reductase (*37*) and human glutathione disulfide reductase (*38*) simply through the two electron transfer from $FADH_2$. But the oxidations of thiol into sulfenic acid, sulfinic acid and sulfonic acid cannot be reversed as easily as the disulfide formation (*32, 39, 40*). Hydrogen peroxide and other biological oxidants such as peroxynitrite can oxidize thiol or thiolate anion into sulfenic acids (*41*). Although this process is also reversible, the electrophilic sulfur atom in sulfenic acid needs to carry out a condensation reaction with a nearby cysteine thiol in the protein or a low-molecular weight thiol such as glutathione to form disulfide, prior to further reduction into thiol (*42*). Unstable and highly reactive sulfenic acid can be readily oxidized into sulfinic acid and sulfonic acid. Sulfinylation was considered irreversible until the discovery of sulfiredoxin, an ATP-dependent enzymatic sulfinic acid reductase (*43*), which specifically reduces the sulfinic acid (*44*) in the peroxiredoxin (*45*). The sulfonylation, however, is irreversible, and the oxidized protein is frequently subjected to ubiquitinaiton and degradation, such as in the case of GTPase-activating proteins (*46*).

In terms of protein function, on one hand, oxidation of the key reactive cysteine thiols can inhibit the function of the redox-sensitive proteins such as glyceraldehyde-3-phosphate dehydrogenase and tyrosine phosphatase (*47*). On the other hand, oxidation can activate the protein in terms of the chaperone function, like in the case of sulfenylation of Hsp33 (*48*) and sulfonylation of peroxiredoxin (*45*). Furthermore, oxidation of cysteine thiol groups into sulfenic acid and sulfinit acid can assist the metal binding, as the example shown in cobalt-containing nitrile reductase

(*49, 50*). Although the detailed process of sulfonylation in THI5p has not been revealed, the sulfonylation in peroxiredoxin has been shown not to be a process of auto-oxidation studied both *in vitro* and *in vivo* (*45*). Moreover, oxidation from sulfinic acid to sulfonic acid in peroxiredoxin has been proved to significantly enhance the chaperone activity from both biochemical data and microscopic imaging, which provides further insights into the enzyme functional change after the irreversible oxidation (*45*).

Despite of the intriguing chemistry that the irreversible oxidation of thiol implies, we cannot rule out the possibility that the modified cysteine could be an artifact during protein crystallization and the pre-treatment for mass spectrometry analysis. In order to further explore the role of cysteine residues and their structural and functional relationship to the iron binding site, UV-Vis, Mössbauer (*19*), electron paramagnetic resonance (*51*) and X-ray absorption spectroscopy (*52*) can be used to elucidate the coordination and oxidation states of iron. Although the type of modification has not been fully determined, all the structural and biochemical evidence indicate that this cysteine residue is highly reactive and can be actively engaged in the reaction.

REFERENCES

1. Weiss, S. (1940) Occidental beriberi with cardiovascular manifestations - Its relation to thiamin deficiency, *J. Am. Med. Assoc. 115*, 832-839.
2. Harper, C. (2006) Thiamine (vitamin B1) deficiency and associated brain damage is still common throughout the world and prevention is simple and safe!, *Eur. J. Neurol. 13*, 1078-1082.
3. Chatterjee, A., Abeydeera, N. D., Bale, S., Pai, P. J., Dorrestein, P. C., Russell, D. H., Ealick, S. E., and Begley, T. P. (2011) Saccharomyces cerevisiae THI4p is a suicide thiamine thiazole synthase, *Nature 478*, 542-546.
4. Chatterjee, A., Li, Y., Zhang, Y., Grove, T. L., Lee, M., Krebs, C., Booker, S. J., Begley, T. P., and Ealick, S. E. (2008) Reconstitution of ThiC in thiamine pyrimidine biosynthesis expands the radical SAM superfamily, *Nat. Chem. Biol. 4*, 758-765.
5. Lai, R. Y., Huang, S., Fenwick, M. K., Hazra, A., Zhang, Y., Rajashankar, K., Philmus, B., Kinsland, C., Sanders, J. M., Ealick, S. E., and Begley, T. P. (2012) Thiamin Pyrimidine Biosynthesis in Candida albicans : A Remarkable Reaction between Histidine and Pyridoxal Phosphate, *J. Am. Chem. Soc*.
6. Coquille, S., Roux, C., Fitzpatrick, T. B., and Thore, S. (2012) The last piece in the vitamin B1 biosynthesis puzzle: Structural and functional insight into yeast HMP-P synthase, *J. Biol. Chem.*
7. Fontecave, M. (2006) Iron-sulfur clusters: ever-expanding roles, *Nat. Chem. Biol. 2*, 171-174.
8. Bale, S., Rajashankar, K. R., Perry, K., Begley, T. P., and Ealick, S. E. (2010) HMP binding protein ThiY and HMP-P synthase THI5 are structural homologues, *Biochemistry 49*, 8929-8936.

9. Tsujimura, M., Odaka, M., Nagashima, S., Yohda, M., and Endo, I. (1996) Photoreactive nitrile hydratase: the photoreaction site is located on the alpha subunit, *J. Biochem. 119*, 407-413.

10. Huang, W., Jia, J., Cummings, J., Nelson, M., Schneider, G., and Lindqvist, Y. (1997) Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold, *Structure 5*, 691-699.

11. Ausubel, F. M., and Brent, F. (1987) *Curr. Protoc. Mol. Biol.*, John Wiley and Sons, New York.

12. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, In *Methods Enzymol.* (Carter, C. W. J., and Sweet, R. M., Eds.), pp 307-326, Academic Press, New York.

13. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination, *Acta Crystallogr. Sect. D-Biol. Crystallogr. 58*, 1948-1954.

14. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283-291.

15. Fukami-Kobayashi, K., Tateno, Y., and Nishikawa, K. (1999) Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history, *J. Mol. Biol. 286*, 279-290.

16. Krissinel, E., and Henrick, K. (2005) Detection of protein assemblies in crystals, In *CompLife* (Berthold, M. R., Glen, R., Diederichs, K., Kohlbacher, O., and Fischer, I., Eds.), pp 163-174.

17. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol. 372*, 774-797.

18. Krissinel, E. (2010) Crystal Contacts as Nature's Docking Solutions, *J. Comput. Chem. 31*, 133-143.

19. Dewitt, J. G., Bentsen, J. G., Rosenzweig, A. C., Hedman, B., Green, J., Pilkington, S., Papaefthymiou, G. C., Dalton, H., Hodgson, K. O., and Lippard, S. J. (1991) X-ray Absorption, Mossbauer, and EPR Studies of the Dinuclear Iron Center in the Hydroxylase Component of Methane Monooxygenase, *J. Am. Chem. Soc. 113*, 9219-9235.

20. Kovaleva, E. G., and Lipscomb, J. D. (2007) Crystal Structures of Fe2+ Dioxygenase Superoxo, Alkylperoxo, and Bound Product Intermediates, *Science 316*, 453-457.

21. Oh, B. H., Pandit, J., Kang, C. H., Nikaido, K., Gokcen, S., Ames, G. F. L., and Kim, S. H. (1993) 3-DIMENSIONAL STRUCTURES OF THE PERIPLASMIC LYSINE ARGININE ORNITHINE-BINDING PROTEIN WITH AND WITHOUT A LIGAND (VOL 268, PG 11348, 1993), *J. Biol. Chem. 268*, 17648-17649.

22. Sharff, A. J., Rodseth, L. E., Spurlino, J. C., and Quiocho, F. A. (1992) Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis, *Biochemistry 31*, 10657-10663.

23. Tam, R., and Saier, M. H. (1993) STRUCTURAL, FUNCTIONAL, AND EVOLUTIONARY RELATIONSHIPS AMONG EXTRACELLULAR SOLUTE-BINDING RECEPTORS OF BACTERIA, *Microbiol. Rev. 57*, 320-346.

24. Quiocho, F. A., and Ledvina, P. S. (1996) Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: Variation of common themes, *Mol. Microbiol. 20*, 17-25.

25. Shi, R., Proteau, A., Wagner, J., Cui, Q., Purisima, E. O., Matte, A., and Cygler, M. (2009) Trapping open and closed forms of FitE—A group III periplasmic binding protein, *Proteins. 75*, 598-609.

26. Felder, C. B., Graul, R. C., Lee, A. Y., Merkle, H. P., and Sadee, W. (1999) The venus flytrap of periplasmic binding proteins: An ancient protein module present in multiple drug receptors, *AAPS Pharmsci. 1*, art. no.-2.

27. Holm, L., Kaarianinen, S., Rosenstrom, P., and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3., *Bioinformatics 24*, 2780-2781.

28. Santini, S., Claverie, J. M., Mouz, N., Rousselle, T., Maza, C., Monchois, V., and Abergel, C. (2011) The conserved Candida albicans CA3427 gene product defines a new family of proteins exhibiting the generic periplasmic binding protein structural fold, *PLoS One 6*, e18528.

29. Beale, J., Lee, S. Y., Iwata, S., and Beis, K. (2010) Structure of the aliphatic sulfonate-binding protein SsuA from Escherichia coli, *Acta Crystallogr. F-Struct. Biol. Cryst. Commun. 66*, 391-396.

30. Koropatkin, N. M., Pakrasi, H. B., and Smith, T. J. (2006) Atomic structure of a nitrate-binding protein crucial for photosynthetic productivity, *Proc. Natl. Acad. Sci. 103*, 9820-9825.

31. Koropatkin, N. M., Koppenaal, D. W., Pakrasi, H. B., and Smith, T. J. (2007) The structure of a cyanobacterial bicarbonate transport protein, CmpA, *J. Biol. Chem. 282*, 2606-2614.

32. Giles, N. M., Watts, A. B., Giles, G. I., Fry, F. H., Littlechild, J. A., and Jacob, C. (2003) Metal and Redox Modulation of Cysteine Protein Function, *J. Chem. & Biol. 10*, 677-693.

33. Saitoh, M., Nishitoh, H., Fujii, M., Takeda, K., Tobiume, K., Sawada, Y., Kawabata, M., Miyazono, K., and Ichijo, H. (1998) Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1, *EMBO J 17*, 2596-2606.

34. Gotoh, Y., and Cooper, J. A. (1998) Reactive oxygen species- and dimerization-induced activation of apoptosis signal-regulating kinase 1 in tumor necrosis factor-alpha signal transduction, *J. Biol. Chem. 273*, 17477-17482.

35. Zheng, M., Aslund, F., and Storz, G. (1998) Activation of the OxyR transcription factor by reversible disulfide bond formation, *Science 279*, 1718-1721.

36. Lee, S. R., Kwon, K. S., Kim, S. R., and Rhee, S. G. (1998) Reversible inactivation of protein-tyrosine phosphatase 1B in A431 cells stimulated with epidermal growth factor, *J. Biol. Chem. 273*, 15366-15372.

37. Mustacich, D., and Powis, G. (2000) Thioredoxin reductase, *Biochem. J. 346 Pt 1*, 1-8.

38. Pai, E. F., and Schulz, G. E. (1983) The catalytic mechanism of glutathione reductase as derived from x-ray diffraction analyses of reaction intermediates, *J. Biol. Chem. 258*, 1752-1757.

39. Harman, L. S., Mottley, C., and Mason, R. P. (1984) Free radical metabolites of L-cysteine oxidation, *J. Biol. Chem. 259*, 5606-5611.

40. Giles, G. I., and Jacob, C. (2002) Reactive sulfur species: an emerging concept in oxidative stress, *Biol. Chem. 383*, 375-388.

41. Nagy, P., and Ashby, M. T. (2007) Reactive sulfur species: kinetics and mechanisms of the oxidation of cysteine by hypohalous acid to give cysteine sulfenic acid, *J. Am. Chem. Soc. 129*, 14082-14091.

42. Reddie, K. G., and Carroll, K. S. (2008) Expanding the functional diversity of proteins through cysteine oxidation, *Curr. Opin. Chem. Biol. 12*, 746-754.

43. Biteau, B., Labarre, J., and Toledano, M. B. (2003) ATP-dependent reduction of cysteine-sulphinic acid by S. cerevisiae sulphiredoxin, *Nature 425*, 980-984.

44. Jonsson, T. J., Johnson, L. C., and Lowther, W. T. (2008) Structure of the sulphiredoxin-peroxiredoxin complex reveals an essential repair embrace, *Nature 451*, 98-101.

45. Lim, J. C., Choi, H. I., Park, Y. S., Nam, H. W., Woo, H. A., Kwon, K. S., Kim, Y. S., Rhee, S. G., Kim, K., and Chae, H. Z. (2008) Irreversible Oxidation of the Active-site Cysteine of Peroxiredoxin to Cysteine Sulfonic Acid for Enhanced Molecular Chaperone Activity, *J. Biol. Chem. 283*, 28873-28880.

46. Tasaki, T., and Kwon, Y. T. (2007) The mammalian N-end rule pathway: new insights into its components and physiological roles, *Trends. Biochem. Sci. 32*, 520-528.

47. D'Autreaux, B., and Toledano, M. B. (2007) ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis, *Nat. Rev. Mol. Cell Biol. 8*, 813-824.

48. Ilbert, M., Horst, J., Ahrens, S., Winter, J., Graf, P. C., Lilie, H., and Jakob, U. (2007) The redox-switch domain of Hsp33 functions as dual stress sensor, *Nat. Struct. Mol. Biol. 14*, 556-563.

49. Zhou, Z. M., Hashimoto, Y., and Kobayashi, M. (2009) Self-subunit Swapping Chaperone Needed for the Maturation of Multimeric Metalloenzyme Nitrile

Hydratase by a Subunit Exchange Mechanism Also Carries Out the Oxidation of the Metal Ligand Cysteine Residues and Insertion of Cobalt, *J. Biol. Chem. 284*, 14930-14938.

50. Arakawa, T., Kawano, Y., Katayama, Y., Nakayama, H., Dohmae, N., Yohda, M., and Odaka, M. (2009) Structural Basis for Catalytic Activation of Thiocyanate Hydrolase Involving Metal-Ligated Cysteine Modification, *J. Am. Chem. Soc. 131*, 14838-14843.

51. Saunders, A. H., Griffiths, A. E., Lee, K. H., Cicchillo, R. M., Tu, L., Stromberg, J. A., Krebs, C., and Booker, S. J. (2008) Characterization of quinolinate synthases from Escherichia coli, Mycobacterium tuberculosis, and Pyrococcus horikoshii indicates that 4Fe-4S clusters are common cofactors throughout this class of enzymes, *Biochemistry 47*, 10999-11012.

52. Lindahl, P. A., Boon Keng, T., and Orme-Johnson, W. H. (1987) EXAFS studies of the nitrogenase iron protein from Azotobactor vinelandii, *Inorganic Chem. 26*, 3912-3916.

CHAPTER 5

CONCLUSION

The work presented here encompasses a wide range of biological problems in terms of protein structure and function. X-ray crystallography coupled with biochemical analysis provides valuable insights into the molecular mechanism of enzyme catalysis. The snapshots captured along the reaction coordinate not only outline the whole picture of the reaction process, but also reveal subtle changes in the structure that catalyzes the chemistry of life. The structural characterization in various metabolic pathways also enables the rational drug design based on the understanding of the catalytic mechanisms.

The structural characterization and mechanistic study on ΨMP glycosidase in Chapter 2 represents a scenario in which structural information becomes essential when multiple mechanisms appear plausible. As the most abundant modification in RNA, Ψ is post-transcriptionally isomerized by Ψ synthase through the cleavage of glycosidic C-N bond, which reconnects the uracil to the ribosyl moiety. Since the mechanism of modification of uridine into Ψ has been extensively studied (*1*), the microscopic reverse of this process thereby stands for one possibility of the Ψ degradation reaction. Similar to the extensively studied N- and O- glycosidic bond cleavage (*2-8*), this process involves an oxocarbenium ion intermediate through a largely dissociative mechanism. Nevertheless, a novel mechanism emerges after several structures of substrate and intermediate complexes of ΨMP glycosidase are determined. A lysine residue is observed to form covalent adduct with the ring-opened ribose ΨMP. Such adduct releases the uracil while retaining the covalent

linkage between the lysine and a ring-opened R5P. LC-MS analysis and site-directed mutagenesis establishes further evidence for the covalent adducts, indicating a novel ribose ring-opening mechanism. Although an alternative mechanism in which the lysine displaces the pyrimidine from the ribose can be found in other enzymes, the lysine residue modeled in ΨMP glycosidase is not positioned for direct base displacement, and the ring-opened ribose ΨMP complex is not consistent with this mechanism. In all, the four unique crystal structures suggest that the reaction utilizes Lys166 to form a covalent adduct to overcome the energy barrier and facilitate the cleavage of the thermodynamically stable C-C glycosidic bond. The crystal structures, in conjugation with mass spectrometry and kinetic analysis demonstrate that ΨMP glycosidase catalyzes the cleavage of the C-C glycosidic bond through a novel ribose ring-opening mechanism.

Another scenario in which protein structure plays a critical role in elucidating the mechanism is represented in the structure determination of eukaryotic thiamin pyrimidine synthase. Thiamin biosynthesis generally involves the separate formation of the thiazole and pyrimidine moieties, which are then condensed to form the thiamin pyrophosphate. In prokaryotes, the enzymes involved in the biosynthetic pathway of thiamin have been extensively studied (*9*). In eukaryotes, while the formation of the thiazole ring has been recently characterized structurally and biochemically (*10*), little is know about the mechanistic details for the pyrimidine synthesis. The mechanism of pyrimidine synthesis thus remains the last piece of puzzle to complete our understanding of the *de novo* biosynthetic pathway of thiamin across all organisms (*11*). Chapter 3 demonstrates the first mechanistic and structural study of THI5p.

Labeling studies and LC-MS analysis have shown that the HMP-P formation utilizes PLP and a histidine residue from the THI5 protein. In *C. albicans*, THI5p contains six histidine residues, two of which are absolutely conserved among all species. Paradoxically, mutation of any of the two histidine residues abolishes the enzyme activity according to the kinetic assays. Therefore, the histidine residue that serves as the donor for HMP-P formation is unable to be identified with only biochemical analysis. Not until the crystal structure was solved, has the identity of the key histidine residue been revealed. The structure of THI5p complex with PLP shows His66 stacked on top of PLP, which is bound via an imine formed with an active site lysine. Together, the biochemical assay and the crystal structure suggest that His66 of the THI5 protein is the histidine source for pyrimidine formation and the pyrimidine synthase is a single-turnover enzyme. In Chapter 4, further insights are provided for the THI5p after running the reaction with PLP and iron. Unusual difference density is observed for Cys199 as part of the CCCFC motif. The trigonal-shaped large electron density on the cysteine side chain indicates a sulfonic group, which is consistent with the LC-MS result. The modified-THI5p also undergoes significant conformational changes in the phosphate binding domain. Given that THI5p belongs to Type II PBPs according to topological arrangement, the conformational change of the modified THI5p differs from the hinge-binding motion for other PBPs upon the binding and releasing of the ligand. Distinct from the Venus fly trap mechanism as previously proposed (*11*), the conformational changes of THI5p involve local movement of the secondary structures throughout the domain. In addition to the modified-THI5p, the structure of C199A/PLP is determined, with high structural similarity to the native

protein. Given both PLP and iron are added to run the reaction during protein purification, the crystal structure shows that the reaction does not proceed and only PLP is bound as in the THI5p/PLP complex. Determined at high resolution, the C199A/PLP shows the previously undetermined C-terminal as an α-helix followed by a loop, with high flexibility indicated by the B-factor. The C-terminal residues cover on top of the CCCFC motif, which is proposed to be the iron binding site. The spatial proximity can thus enable the C-terminal residues to modulate the dielectric constant around the iron. In order to further investigate the effect of the flexible C-terminus, we mutated THI5p into a C-terminus truncated THI5p, E309Ter. Given similar treatment to the modified-THI5p, with iron and PLP, the structure of E309Ter closely resembles that of the THI5p/PLP complex, in which only PLP is bound and no reaction observed. The structure similarity between E309Ter/PLP and THI5p/PLP implies that the C-terminus does not affect the overall structure of the protein, but could be potentially involved in the iron binding.

Each of the structure above represents the functional and mechanistic insights that structural biology has brought us to the understanding of life science. The structure of ΨMP glycosidase reveals, for the first time, the mechanism of the C-C glycosidic bond cleavage in nature. The structure of THI5p demonstrates its distinct character as a suicide enzyme and has intrigued active research in the mechanistic details of the complicated chemistry. Given the tremendous effort to elucidate the structural details of these proteins, more fascinating research and collaborations are going on across the field of biology, chemistry and physics in this new era of structure biology.

REFERENCES

1. Miracco, E. J., and Mueller, E. G. (2011) The products of 5-fluorouridine by the action of the pseudouridine synthase TruB disfavor one mechanism and suggest another, *J. Am. Chem. Soc. 133*, 11826-11829.
2. Gopaul, D. N., Meyer, S. L., Degano, M., Sacchettini, J. C., and Schramm, V. L. (1996) Inosine-uridine nucleoside hydrolase from *Crithidia fasciculata*. Genetic characterization, crystallization, and identification of histidine 241 as a catalytic site residue, *Biochemistry 35*, 5963-5970.
3. Duerre, J. A. (1962) Hydrolytic Nucleosidase Acting on *S*-Adenosylhomocysteine and on 5'-Methylthioadenosine, *J. Biol. Chem. 237*, 3737-&.
4. Hurwitz, J., Heppel, L. A., and Horecker, B. L. (1957) The Enzymatic Cleavage of Adenylic Acid to Adenine and Ribose 5-Phosphate, *J. Biol. Chem. 226*, 525-540.
5. Pugmire, M. J., and Ealick, S. E. (1998) The crystal structure of pyrimidine nucleoside phosphorylase in a closed conformation, *Structure 6*, 1467-1479.
6. Flaks, J. G., Erwin, M. J., and Buchanan, J. M. (1957) Biosynthesis of the Purines .16. the Synthesis of Adenosine 5'-Phosphate and 5-Amino-4-Imidazolecarboxamide Ribotide by a Nucleotide Pyrophosphorylase, *J. Biol. Chem. 228*, 201-213.
7. Olsen, A. S., and Milman, G. (1974) Chinese-Hamster Hypoxanthine-Guanine Phosphoribosyltransferase - Purification, Structural, and Catalytic Properties, *J. Biol. Chem. 249*, 4030-4037.
8. Short, S. A., Armstrong, S. R., Ealick, S. E., and Porter, D. J. T. (1996) Active site amino acids that participate in the catalytic mechanism of nucleoside 2'-deoxyribosyltransferase, *J. Biol. Chem. 271*, 4978-4987.

9. Jurgenson, C. T., Begley, T. P., and Ealick, S. E. (2009) The structural and biochemical foundations of thiamin biosynthesis, *Annu Rev Biochem 78*, 569-603.

10. Chatterjee, A., Abeydeera, N. D., Bale, S., Pai, P. J., Dorrestein, P. C., Russell, D. H., Ealick, S. E., and Begley, T. P. (2011) Saccharomyces cerevisiae THI4p is a suicide thiamine thiazole synthase, *Nature 478*, 542-546.

11. Coquille, S., Roux, C., Fitzpatrick, T. B., and Thore, S. (2012) The last piece in the vitamin B1 biosynthesis puzzle: Structural and functional insight into yeast HMP-P synthase, *J. Biol. Chem.*